AFRL-OSR-VA-TR-2015-0167

A New Paradigm to Identify Reaction Pathways in Gas-phase

Angela Violi
UNIVERSITY OF MICHIGAN

04/29/2015
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTE
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
27-04-2015	Final Report	01 FEB 2013 - 30 JAN 2015

4. TITLE AND SUBTITLE
A New Paradigm to Identify Reaction Pathways in Gas-phase

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-13-1-0031

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Angela Violi and Paolo Elvati

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
REGENTS OF THE UNIVERSITY OF MICHIGAN
DIVISION OF RESEARCH DEVELOPMENT AND ADMINISTRATION
503 THOMPSON ST ANN ARBOR MI 48109-1340

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 N. Randolph Street
Suite 325 Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The complexity of the energy landscapes of hydrocarbon molecules during combustion processes, often composed by several hundreds of minima and even more barriers, makes the heuristic search of the most likely reactions an unfeasible task even with today's computer power. As a consequence, the main advancements in the field of combustion chemistry and kinetic mechanisms development are based on "chemical intuition" and trial and error procedures, which are error prone and hard to automate. In this proposal we present a new paradigm to determine reaction pathways for gas-phase species that relies on two major components: molecular dynamics simulations and advanced sampling techniques. Molecular dynamics (MD) in conjunction with advanced sampling techniques, such as Metadynamics, are used to explore the energy landscapes of uni-molecular and bi-molecular reactions in gas phase. Starting from the fuel molecule and running several simulations from the various wells identified with MD, we can recover a network of reactions that includes a controlled number of reactions, allowing the construction of reaction pathways with the desired level of detail.

15. SUBJECT TERMS
Reaction pathways, combustion kinetics.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Chiping Li
Unclassified	Unclassified	Unclassified		1	19b. TELEPHONE NUMBER *(Include area code)* 703-696-8574

Reset

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

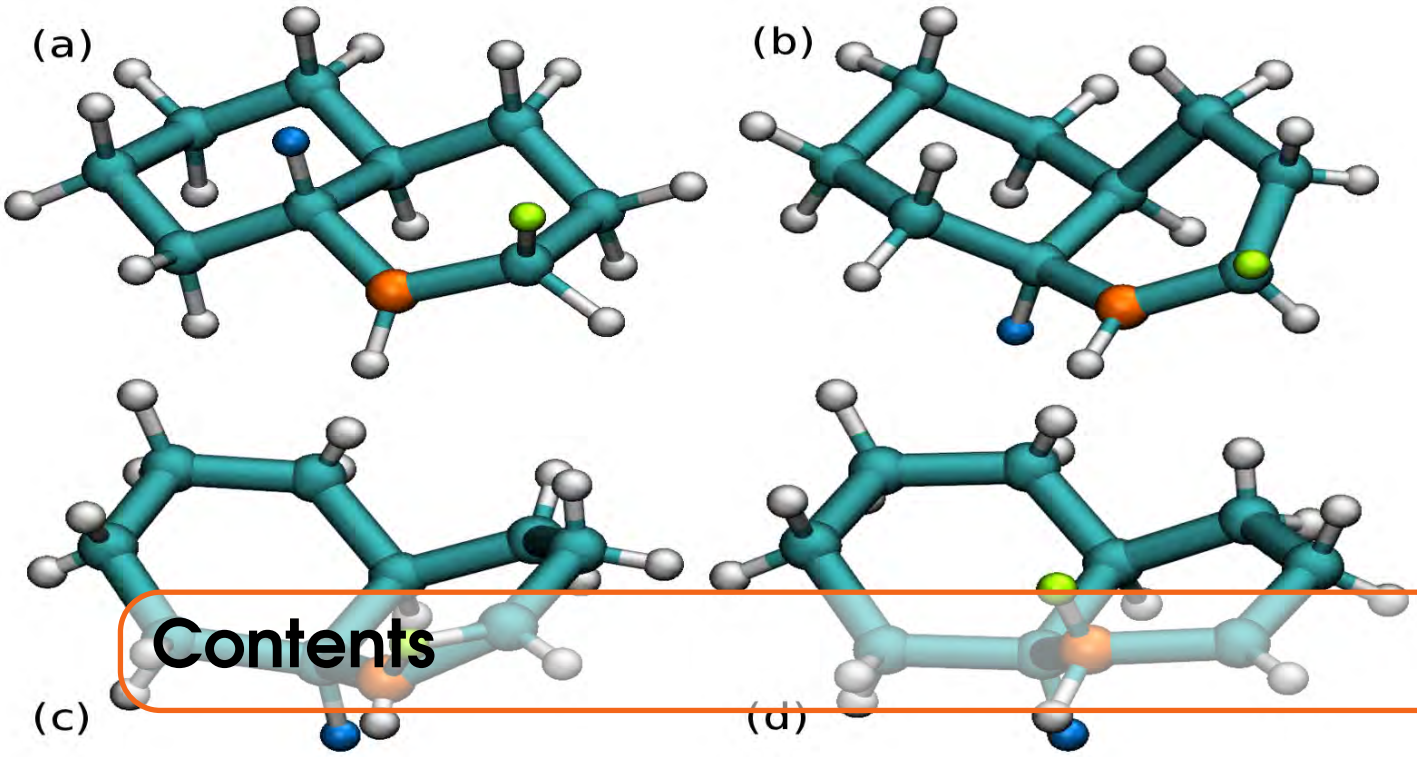
A New Paradigm to Identify Reaction Pathways in Gas-phase

F.E.R.N.

Fast Exploration of chemical Reaction Networks

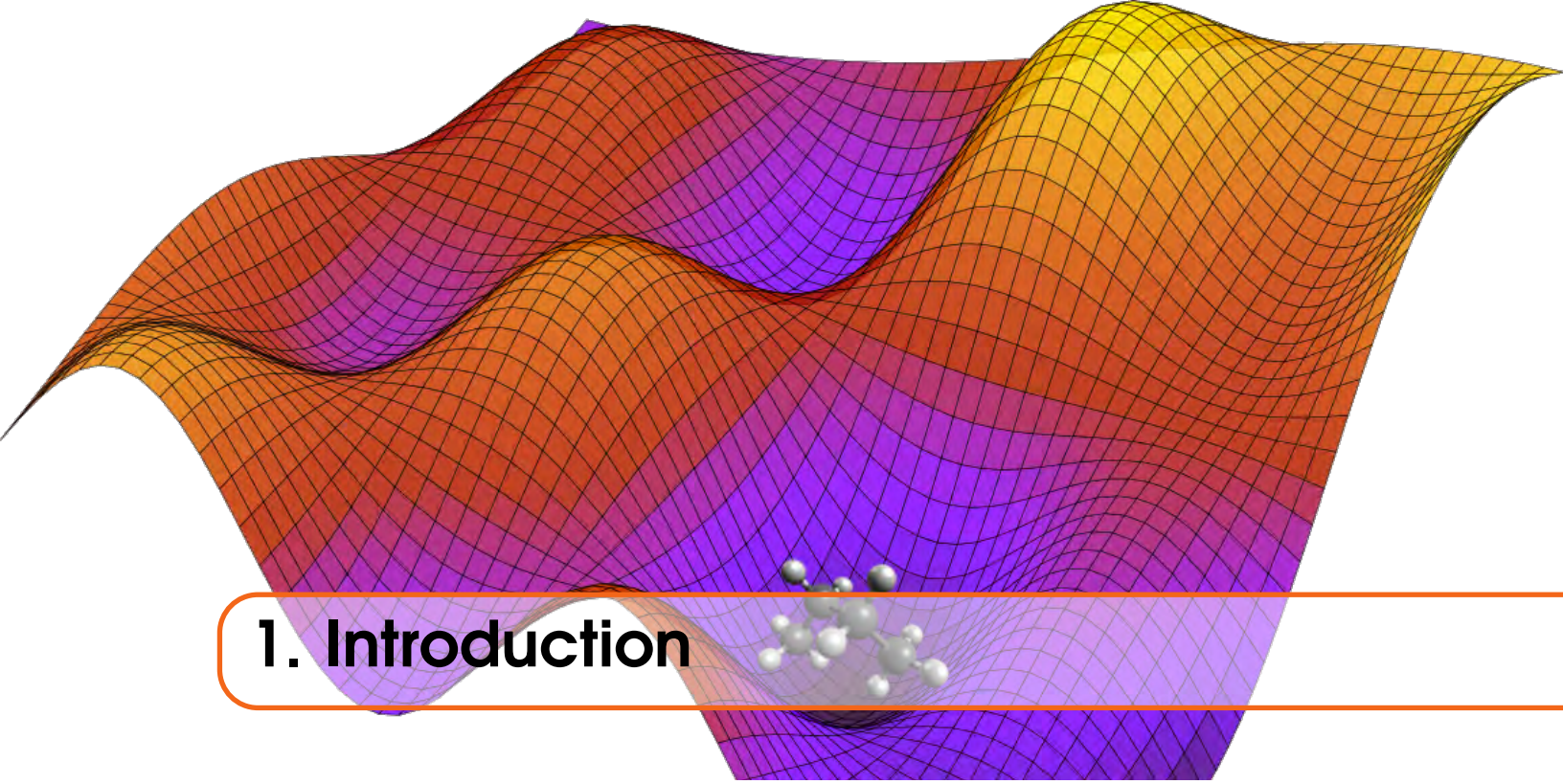
Angela Violi, Paolo Elvati

This research is funded by the U.S. Air Force Office of Scientific Research (Grant FA9550-13-1-0031) under the technical supervision of Dr. Chiping Li.
Last edited: 2015-04-23



Contents

Foundations



1. Introduction

Many industrial and natural processes, like nucleation or combustion are the result of thousands of interdependent reactions. Seemingly, a model of these processes demands all these reactions to be taken into account without regard to their ultimate importance. However, despite the fast growth of computational resources of the last few decades, this extensive modeling remains a daunting task except when massive simplifications are made. This situation mostly originates from two problems: high computational cost of the methods that need to be employed to accurately model each reaction [1], and from the difficulty of finding and systematically exploring all the possible reactive pathways. Moreover, even if issues related to computational cost will be resolved, the amount of information needed to fully describe an extensive reaction network, would still not help in the understanding of the underlying process as such load of information would be difficult to interpret.

Luckily, in many cases not all the possible reactions have the same importance, as often many reaction pathways contribute only marginally to the products' formation. This fact has been leveraged in the past to build simplified models, often called reduced mechanism (RM), which capture the key aspects of a more detailed description while making it more tractable [2]. A RM is built on two distinct components, a subset of reactions and their corresponding rates; it is generally constructed through a top-down approach, where the complete reaction network is reduced to a more manageable subset [3]. However except for the most simple cases, key reactions are difficult to identify and important simplifications are made in the calculations of the rates [4], which contrasts with the high-level accuracy required to correctly model chemical reactions [5, 6]. Analogy and intuition can be used to guess which are the most critical reactions ans species in a complex reaction network, but even though these ideas are then verified by rigorous testing, this approach is inefficient as it relies on scientific and

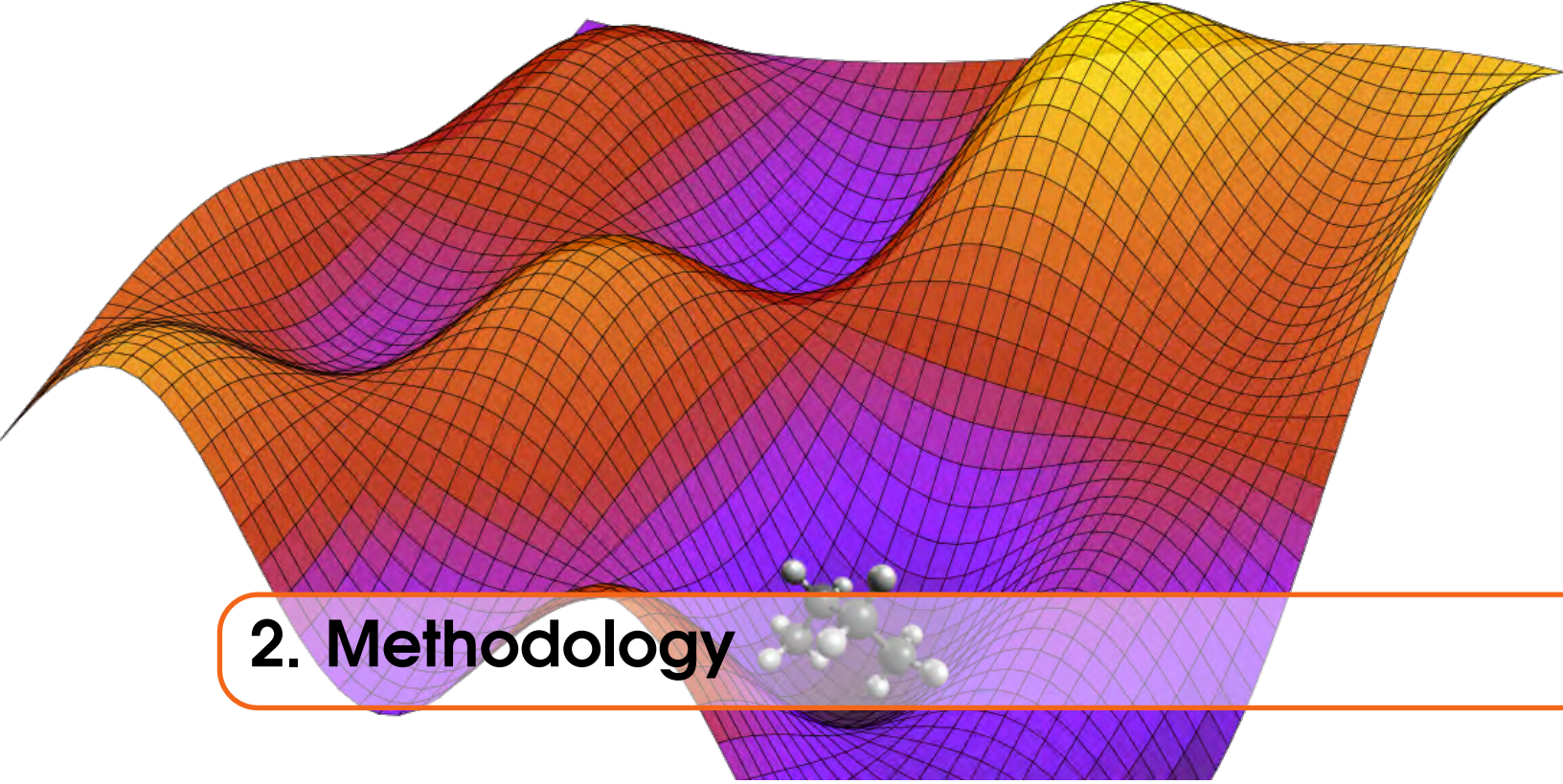
personal bias.

The high sensitivity of the RM to the methods used to compute the rates k stems from the exponential dependency from the free energy (FE) difference between products and transition state $\Delta G^{\ddagger}$ [7]:

$$k = \kappa \frac{k_B T}{h} \exp(-\beta \Delta G^{\ddagger}), \tag{1.1}$$

where κ is the transmission coefficient, k_B is the Boltzmann constant, T is the temperature, h is the Planck's constant and $\beta = (k_B T)^{-1}$. From Eq. 1.1 follows that any error in computing $\Delta G^{\ddagger}$ is likely to have important effects on the values of the derived rates, which may have catastrophic repercussions if the error affects an early branching of a reaction network.

For these reasons, instead of using one of the several existing methods developed to compute the reaction dynamics [8, 9, 10, 11, 12] on all the possible reactions, we developed a new technique to identify and select the most frequent pathways of a reaction network with minimal computational effort by using an acceleration-detection scheme. Briefly, this new approach finds the critical reactions of the RM gradually by iteratively identifying the most common products of given reactants. At each iteration, the products of the previous step form the pool for the reactants together with the most common gas species present under specific conditions. For given reactants, the most frequent reactions are obtained by accelerating their dynamics by repeatedly adding small amount of bias to the FE hyper-surface, in a manner that emulates the energy transfer associated with gas phase collisions. As a result a list of the primary pathways is produced; the corresponding rates can then to be computed with accurate *ab-initio* techniques in order to build the final RM.



2. Methodology

2.1 Overview

For the reasons described in the introductions, instead of using one of the several existing methods developed to compute the reaction dynamics [8, 9, 10, 11, 12] on all the possible reactions, we developed a new technique to identify and select the most frequent pathways of a reaction network with minimal computational effort, leaving the a more accurate estimation of the rates of the selected reactions to high level *ab-initio* calculations. To develop our approach, we started from the observation that for any given reactant(s), all the reaction pathways and rates can be recovered by simply observing the behavior of a large number of replicas of the same system for a long time and counting the occurrences of each reaction. However, this method is not practical because it requires for each reactant several hundred very long simulations, due to the high energy barriers commonly involved in chemical reactions. To make this idea applicable, we employed an acceleration-detection scheme, where the dynamics of all the system's replicas are accelerated until a reaction is detected. The simulations are then interrupted and, after all the replicas are terminated, the frequency of each reactive pathway is calculated. This method has several advantages:

- it does not require *a priori* knowledge of the pathways or transition states;
- it does not rely on the life time or stability of the products, as the simulations are interrupted as soon as the reaction happens, and therefore this approach can also handle pathways where chemical activation plays an important role.

For given reactant(s) this approach will produce a list of pathways and by repeating the same procedure only for the most frequently observed reactions, the key pathways of the entire reaction network are obtained without the need to either map the complete reaction network or

arbitrarily select pathways.

2.2 Acceleration

Since the idea of accelerating systems to overcome energetic barriers is not new [13], we used the basic algorithm that is use in Metadynamics (META) [14], an already well-tested method [15] that uses a history-dependent bias to favor the exploration of new states. Briefly, the well-tempered Metadynamics (WTM) technique was introduced to reconstruct the FE landscape of a system projected on a few coordinates, called collective variables (CVs), by adding an history dependent bias potential that favors the exploration of configuration that are far from equilibrium. Moreover, to get the added potential to convergence to the underlying FE, in the WTM algorithm the amount of added bias is gradually reduced in previously visited values of the CVs. The amount of this reduction is controlled by the bias factor (BF), γ

$$\gamma = \frac{T + \Delta T}{T} \tag{2.1}$$

where T is the temperature of the system and $T + \Delta T$ is the fictitious temperature of the CVs, as in the long time limit the probability distribution of the CVs is

$$P(\mathrm{CV}) \propto e^{-\frac{\mathrm{FE}}{\gamma k_b T}} \tag{2.2}$$

The reader can find a more complete description of META and WTM in the cited papers.

Even though META was originally introduced to reconstruct FE landscapes, here we employ *it only to accelerate the system dynamics*, simplifying the requirements on the definition of the CV. In particular, we use the algorithm to bias the potential energy, forcing the reacting molecule to experience energy fluctuations typical of higher temperatures [16]. This approach can be viewed as a way to force the molecule under investigation to experience several collisions with virtual particles that only increase its internal energy, effectively accelerating its reactions in the high-pressure regime.

With the addition of the potential energy bias Φ, the effective FE experienced by the system becomes $\Delta G_j^{\ddagger} + \Phi$, which substituted in Eq. 1.1 can be used to compute the probability p_i to observe a specific reaction i:

$$p_i = \frac{\kappa_i M_i \exp(-\beta \Delta G_i^{\ddagger})}{\sum_j \kappa_j M_j \exp(-\beta \Delta G_j^{\ddagger})} \tag{2.3}$$

where M_i is the pathway degeneracy of the i-th reaction, and the summation is performed over all the possible reactions.

While biasing the potential energy is an effective and general way to accelerate the system reactivity [17], at the same time it is not an appropriate quantity to use to monitor the evolution of the reactions, since in many cases it is unable to distinguish between products and reactants. A more apt choice for the reaction detection is the measure of the molecular connectivity, like the recently introduced Social PeRmutation INvarianT coordinates (SPRINT) [18]. This class of reaction coordinates defines the connectivity of a system of N atoms with a N-dimensional

vector by using spectral graph theory and including both local and long-range system topology information. SPRINT coordinates have already been successfully used to differentiate and cluster molecular structures [19], by considering the evolution of the Euclidean norm of the difference between instantaneous and average value of the SPRINT.

2.3 SPRINT

The SPRINT CVs are related to the concept of coordination number (CN) of each atom. As the CN is so central for the detection part of the algorithm to work it is worth tune its parameters so that they are optimized for the Fast Exploration of Reaction Network (FERN) algorithm. Due to the derivability requirements off all the CVs in the META technique, a smooth asymmetric step function is needed. We employed the one described by the following equation ($m > n$):

$$cn(r) = \begin{cases} 1 & r \leq d_0 \\ \dfrac{1-(\frac{r-d_0}{r_0})^n}{1-(\frac{r-d_0}{r_0})^m} & r > d_0 \end{cases} \tag{2.4}$$

To define this function four parameters needs to be provided: d_0, r_0, n, and m. While d_0 simply shifts the curve on the r axis, and represents the maximum distance for which a specific CN is always equal to one, the role of the other parameters is slightly more complex. To understand their effect we can consider the three scenarios:

$$r \to d_0 \qquad cn \approx \frac{1-r}{1-r^m} \tag{2.5}$$

$$r \to \inf \qquad cn \approx r^{n-m} \tag{2.6}$$

$$r = d_0 + r_0 \qquad cn = n/m \tag{2.7}$$

From equation 2.5 we can see that the behavior a short distance is dictated by m, while from equation 2.6 can be deduced that in order to have a slow decaying function we need to keep $m-n$ small; finally the third case (equation 2.7) shows that r_0 measures the interval required to go from 1 to n/m.

The exact values to assign to these variables depends on the goal of the simulations. As we are interested in bond breaking, the requirements that we want to meet are:

- minimize the effect on SPRINT of thermal oscillation for a given configuration.
- avoid the collapse of SPRINT values to 0 that follows excessive separation of atoms. Is worth nothing that this requirement is only relatively important since each reaction is interrupted as soon as a clear change in the connectivity is observed and therefore, a weak connectivity even after a bond is broken is not strictly required.
- assign the same (integer) value of n and m among all the all the types of bonds, due to programming restraints. While this requirement can be remove by rewriting the implementation of the SPRINT calculation, it was not removed as it is of no influence on the results.

With these requirements in mind, we first chose $n = 2$ and $m = 3$, which give the least steep function with the slowest decay ($n = 1$ causes a discontinuity in the first derivative at d_0). Then we fine tuned the values of d_0 and r_0 for each bond type by computing the FE as a function of the bond length. As our test are performed on hydrocarbons (see next chapter), here we report the results for the C−C, C−H, and H−H bonds, but the procedure is equivalent for other type of atoms.

2.3.1 C−C, C−H, and H−H bond analysis

The FE were computed by using WTM biased molecular dynamics (MD) simulations performed with the LAMMPS [20] software coupled with the PLUMED [21] plugin (version 1.3). The reaction were carried by employing the AIREBO [22] classic reactive force field (FF). The equations of motion were integrated with a timestep of 0.1 fs; systems were simulated for 50 ns, at a temperature of 2500 K maintained with a Langevin thermostat [23] with a time constant $\tau_{Lan} = 5$ fs; the cutoff distance multiplier was set to 3. As the Gaussian shaped bias with an initial height h of 0.02 eV and a width σ of 0.025 nm was deposited every 50 fs on the specific bond distance while employing a BF of 6. The bond exploration of the phase space was controlled by placing an harmonic "wall" at a distance of 0.5 nm with an elastic constant of 50 eV/nm^2

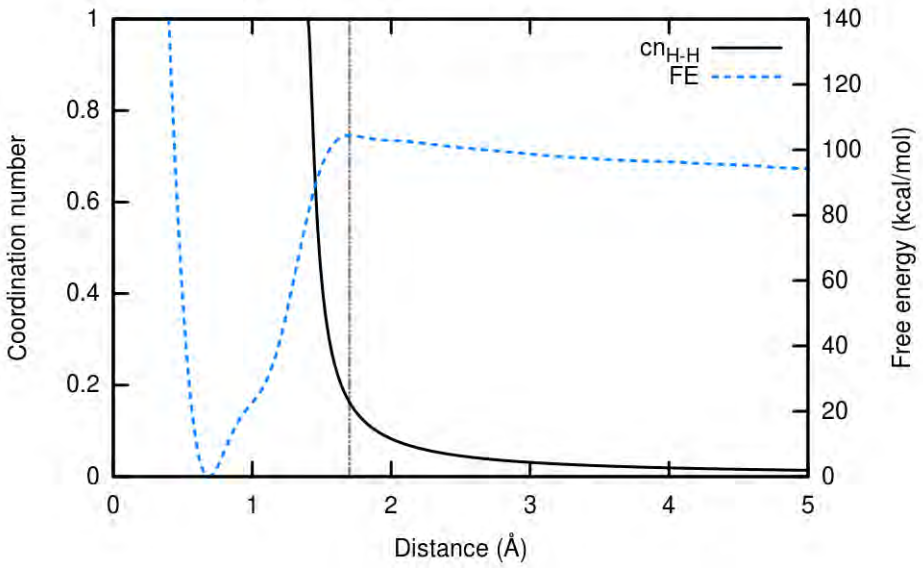


Figure 2.1: (right axis) FE as a function of the H−H bond length in H_2. (left axis) CN as a function of H−H distance. The gray vertical line indicates the approximate location of the transition state.

Due to the cross-dependence of the parameters for the different bonds on the value of the SPRINT we first defined the CN parameters for H−H bond by studying H_2 (Figure 2.1), we then moved to the C−H bond by looking into CH_4 (Figure 2.2), and finally on the C−C bond by analyzing n-C_7H_{16} (Figure 2.2).

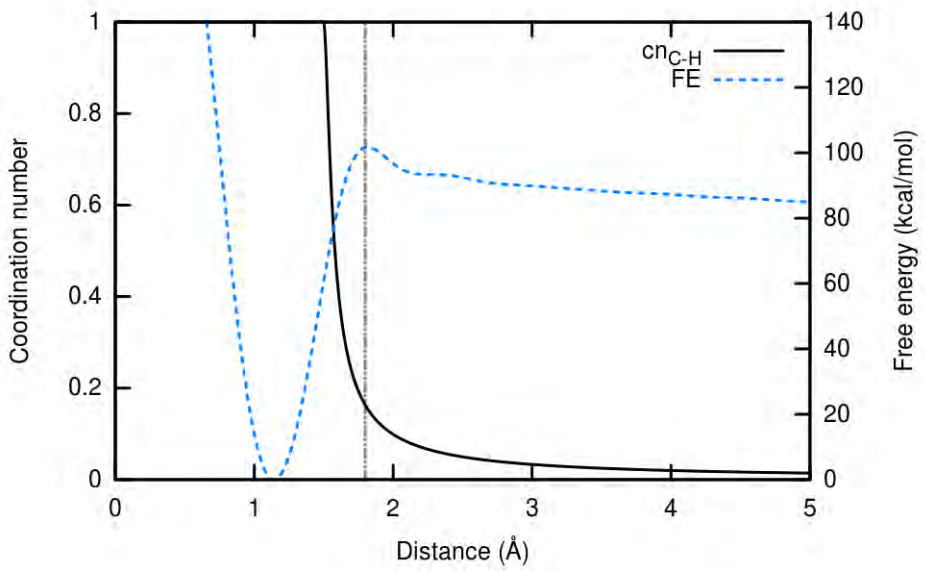


Figure 2.2: (right axis) FE as a function of the C−H bond length in CH_4. (left axis) CN as a function of C−H distance. The gray vertical line indicates the approximate location of the transition state.

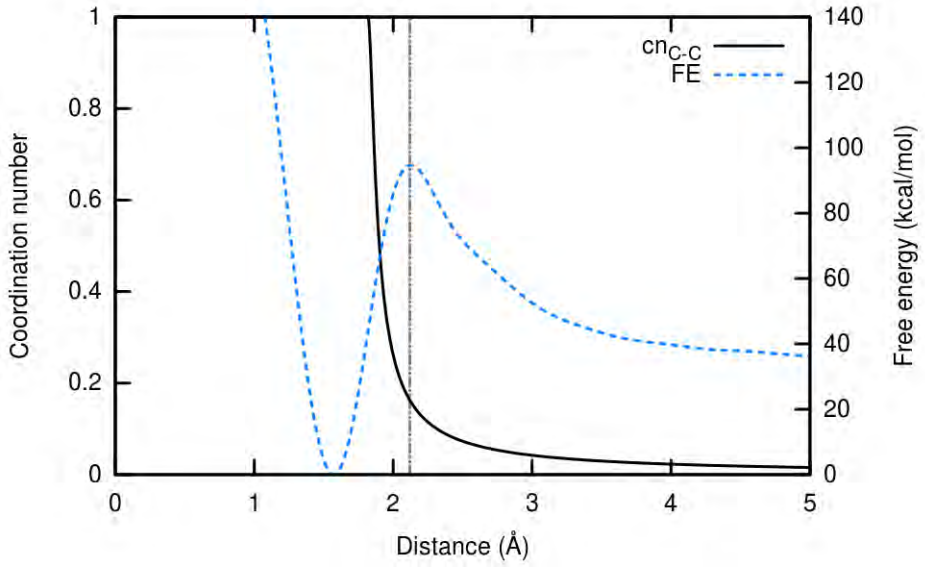
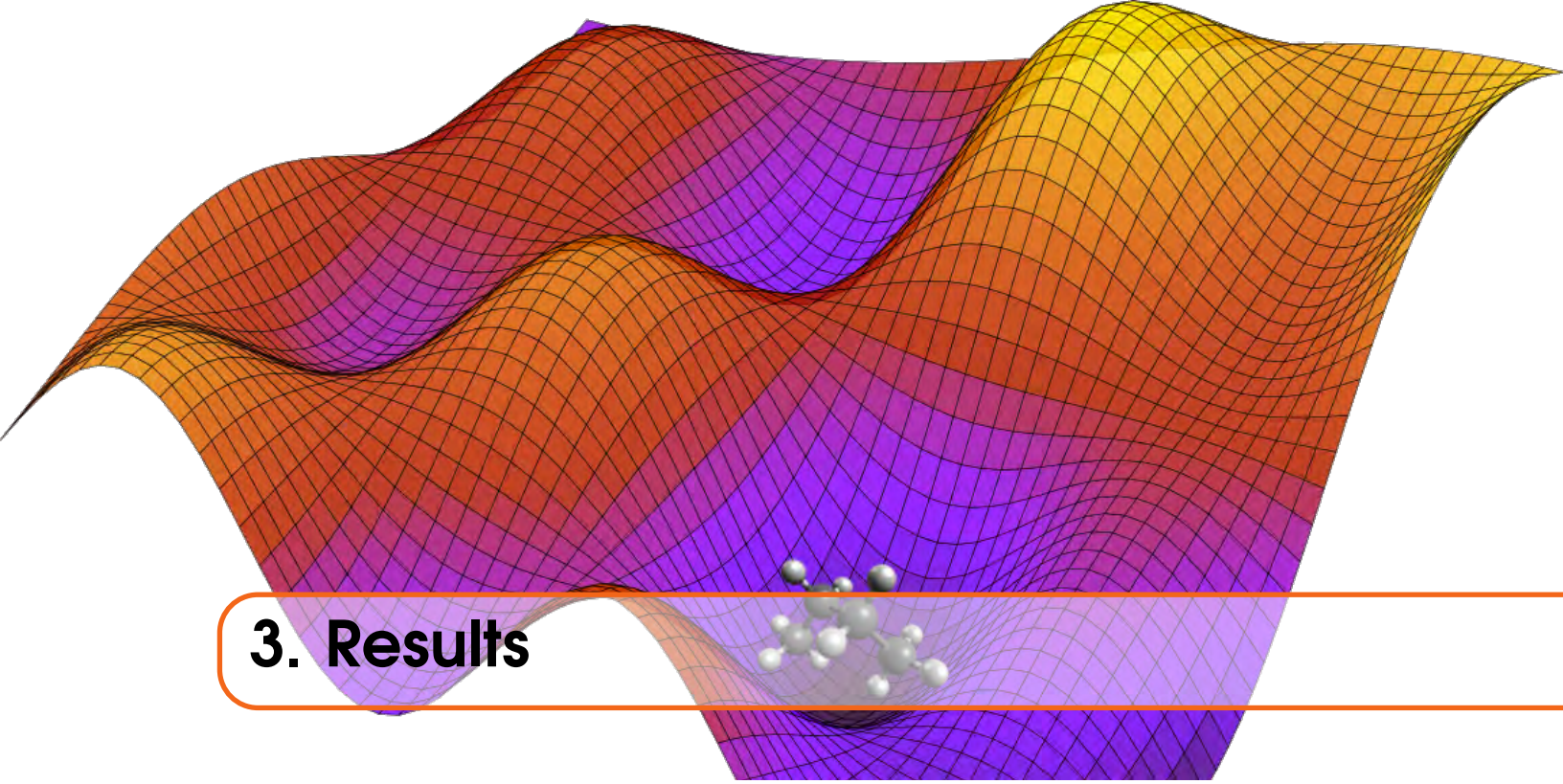


Figure 2.3: (right axis) FE as a function of the C−C bond length of two terminal carbons in *n*-heptane. (left axis) CN as a function of C−C distance. The gray vertical line indicates the approximate location of the transition state.

By analyzing the FES, we found that by assigning $r_0 = 0.05$ and a value to d_0 0.03 nm smaller than the position of the transition state (shown with a gray vertical line in the figure), allows to detect only configuration which have enough energy to potentially cross the FE

barrier, while at the same time ignoring "unproductive" thermal fluctuations even at high temperature. The final values of d_0 for the H−H, C−H, and C−C bonds are 0.14 nm, 0.15 nm, and 0.182 nm, respectively.

Results



3. Results

To illustrate the salient details of our acceleration-detection approach, we applied our methodology to three different systems of increasing complexity. **First**, we studied the behavior of a monatomic particle subjected to an analytical two dimensional FE potential. This is a simple system for which all the properties are known exactly and therefore can be used to test the correctness of the method's foundations. The **second** case we considered is the first step of the unimolecular dissociation of ethane. Compared to the first, this system has the added complexity of the rotational and vibrational modes of the molecules, so we could test and discuss all the issues related to the thermostatting and temperature control. Moreover, as we are employing a polyatomic systems, we illustrate the how to set up the parameters used to compute the SPRINT. The **third** and last system, we studied is the reactivity of *t*-decalin with methyl radical. Not only the number of pathways for this system is larger, but also details on bimolecular reactions are presented.

3.1 Monoatomic particle in 2D potential

The behavior of a single particle in a customizable analytic potential is an excellent test for the performance of the FERN method. The potential was shaped to reproduce three minima separated by two barriers (symmetric or asymmetric, depending on the values of the chosen parameters), as shown in Figure 3.1.

The potentials are described by the following equation:

$$U(x,y) = a_x x^6 + 7a_x x^4 + 12a_x x^2 + d_x x + a_y y^2 \tag{3.1}$$

where d_x can be used to change the difference in energy between the barriers. All the

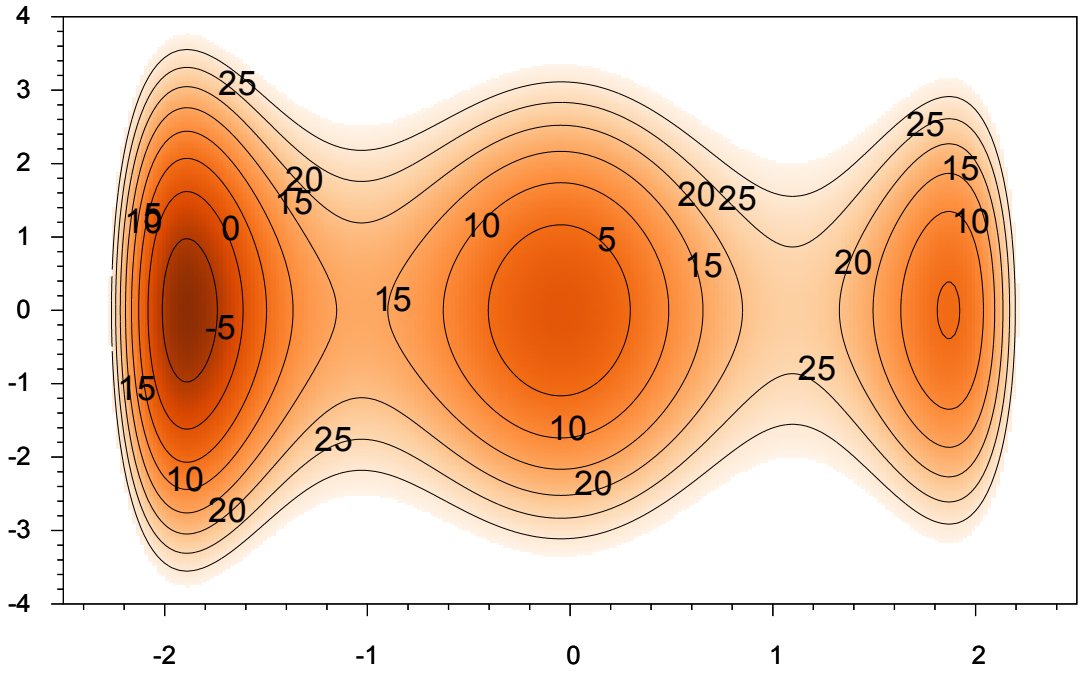


Figure 3.1: Example of the two dimensional potential experienced by the single atom ($a_x = 3$, $d_x = 3.3$, $a_y = 3$). The energy is expressed in kcal/mol and the contour lines are separated by 5 energy units.

simulations were started from the central basin and the rates of the formation of "products" were computed by monitoring the position of the particles on the x axis, for in this simple case there is no chemical connectivity that justifies the use of SPRINT. The simulations were as soon as the atom entered either the negative ($x < -1$) or the positive ($x > 1$) basins.

By tweaking the parameters that define the potential, we created several scenarios with different barrier heights (from about 4 to 80 times k_BT) and different degrees of asymmetry between the negative and positive basins (with a difference between the two barriers ranging from 0 to about 15 k_BT). We analyzed the effect of the thermostat and a variety of META parameters, in particular deposition rate (δ_H), bias shape (height h_H and width σ_H of the Gaussian shaped bias) and BF, when employing the well-tempered version [11].

Initial tests were performed on symmetric systems ($d_x = 0$) with (Table 3.2) and without (Table 3.1) a thermostat. Both WTM, with different values of the BF and "plain" META (BF $= \infty$) were employed. For each system we reported the average and standard deviation (over all the runs) of the temperature($\langle T \rangle$ and σ_T, both in K), and of the simulation time ($\langle t \rangle$ and σ_t, both in ps).

In these tests we expect to statistically observe the atom end in the positive and negative well the same number of time, therefore these tests can be used to evaluate the dependency of the method on different simulations parameters that have nothing to do with the underlying physical system. In all cases, the discrepancy between the observed and the expected probability are within a 95% confidence interval. Interestingly however, increasing the amount of bias

Table 3.1: Results for different symmetric ($d_x = 0$) two dimensional potentials systems simulated without a thermostat. Bias shape parameters (h_H, σ_H) as well as FE barriers ($\Delta G^{\ddagger}_{CC}$, $\Delta G^{\ddagger}_{CH}$) are expressed in kcal/mol, while bias addition frequency δ_H is expressed in ps; masses are listed in Dalton. Temperature average and standard deviation ($\langle T \rangle$, σ_T) are expressed in K; time average and standard deviation ($\langle t \rangle$, σ_t) are in ps. The "Neg." and "Pos." columns refer to the number of observed simulation ending in the negative ($x < -1$) or the positive ($x > 1$) basins, respectively.

BF	h_H	σ_H	δ_H	a_x	a_y	Runs	$\langle T \rangle$	σ_T	$\langle t \rangle$	σ_t	Neg.	Pos.
∞	0.05	0.05	1	2	2	60	639	300	2.19	2.05	31	29
3	0.05	0.05	1	2	2	60	570	246	2.77	2.89	32	28
3	0.01	0.01	1	2	2	60	1017	843	2.08	1.87	28	32
3	0.01	0.01	0.1	2	2	60	1004	288	1.01	1.31	35	25
3	0.01	0.01	0.05	2	2	60	965	336	0.74	0.69	29	31
3	0.1	0.1	10	2	2	60	301	112	24.10	22.55	25	35
3	0.01	0.03	0.05	2	2	60	867	255	1.58	1.38	34	26
3	0.01	0.1	0.05	2	2	60	403	155	10.63	4.82	26	34
10	0.01	0.1	0.05	3	3	100	1250	420	2.17	2.34	47	53
3	0.05	0.05	1	4	2	60	1286	466	2.39	2.78	32	28
3	0.05	0.05	1	4	4	60	1341	541	1.91	1.93	27	33
10	0.01	0.1	0.05	10	2	100	3587	1237	2.09	2.61	58	42
3	0.01	0.1	0.05	10	10	160	3507	1113	2.36	1.90	83	77
10	0.01	0.1	0.05	10	10	160	3823	1372	2.70	2.28	79	81
25	0.01	0.1	0.05	10	10	160	3980	1264	1.65	2.26	89	71
50	0.01	0.1	0.05	10	10	160	3724	1203	1.90	2.28	90	70
100	0.01	0.1	0.05	10	10	160	3904	1392	1.80	2.12	86	74

deposited at each interval does not lead to shorter simulations. As expected in the simulations with the thermostat, we found that very small τ are almost never beneficial, slowing the diffusion of the atom on the FE surface.

A second set of tests was performed with asymmetric potentials, again with (Table 3.3) and without (Table 3.4) the presence of a thermostat. As in the previous case, the application of a thermostat is expected not to affect the ratio between the rates due to the lack of vibrational degrees of freedom of the system. The $\Delta G^{\ddagger}$ needed to compute the theoretical ratio was evaluated directly from 3.1, since there is no relevant difference in the entropic contribution of the two pathways.

A subset of the results (selected for clarity) is shown in Figure 3.2.

In all cases we found an excellent agreement between the pathway probability computed with our approach and the one predicted by using the analytical values of $\Delta G^{\ddagger}$ in Equation 2.3.

Table 3.2: Results for different symmetric two dimensional potential ($a_x = 4$, $a_y = 2$, $d_x = 0$) systems simulated with a thermostat. Bias shape parameters (h_H, σ_H) as well as FE barriers ($\Delta G^{\ddagger}_{CC}$, $\Delta G^{\ddagger}_{CH}$) are expressed in kcal/mol, while bias addition frequency δ_H is expressed in ps; masses are listed in Dalton. Temperature average and standard deviation ($\langle T \rangle$, σ_T) are expressed in K; time average and standard deviation ($\langle t \rangle$, σ_t) are in ps. The "Thermo" column refers to the type of thermostat (Lan for Langevin and NHc10 for a Nose-Hoover chain of length 10), applied with a time constant τ (in ps). The "Neg." and "Pos." columns refer to the number of observed simulation ending in the negative ($x < -1$) or the positive ($x > 1$) basins, respectively.

BF	h_H	σ_H	δ_H	Thermo	τ	Runs	$\langle T \rangle$	σ_T	$\langle t \rangle$	σ_t	Neg.	Pos.
3	0.05	0.05	1	Lan	10	60	1462	427	1.95	2.72	35	25
3	0.05	0.05	1	Lan	1	60	1575	430	1.04	0.72	34	26
3	0.05	0.05	1	Lan	0.1	58	1944	644	3.86	9.48	30	28
3	0.05	0.05	1	NHc10	10	60	1518	425	1.02	2.04	26	34
3	0.05	0.05	1	NHc10	1	59	1535	550	1.57	6.44	26	33
3	0.05	0.05	1	NHc10	0.1	60	1453	451	7.22	12.56	31	29
3	0.01	0.1	0.05	NHc10	0.1	38	901	128	21.30	11.61	21	17
3	0.01	0.1	0.05	NHc10	10	100	897	440	4.66	3.82	45	55

Table 3.3: Results for different asymmetric two dimensional potentials systems simulated with a thermostat. The FE barriers ($\Delta G^{\ddagger}_{CC}$, $\Delta G^{\ddagger}_{CH}$) are expressed in kcal/mol, while masses are listed in Dalton. The bias ($h_H = 0.01$ kcal/mol, $\sigma_H = 0.1$ kcal/mol) was added with a frequency δ_H of 0.05 ps. Temperature average and standard deviation ($\langle T \rangle$, σ_T) are expressed in K; time average and standard deviation ($\langle t \rangle$, σ_t) are in ps. τ (in ps) is the time constant of the Nose-Hoover chain thermostat (length 10) "Neg." and "Pos." columns refer to the number of observed simulation ending in the negative ($x < -1$) or the positive ($x > 1$) basins, respectively.

BF	a_x	a_y	d_x	τ	Runs	$\langle T \rangle$	σ_T	$\langle t \rangle$	σ_t	Neg.	Pos.
3	2	2	0.6	0.01	120	508	19	19.87	14.24	107	13
3	2	2	0.6	0.05	52	567	61	7.27	3.77	46	6
3	2	2	0.6	0.1	98	656	153	7.06	10.86	86	12
3	2	2	0.6	0.05	100	1389	432	1.17	2.86	64	36
1.5	2	2	0.6	0.05	97	1228	275	3.52	3.91	74	23
50	2	2	0.6	0.05	100	763	215	3.29	5.67	83	17
50	8	8	0.15	0.05	65	1657	204	17.95	5.09	47	18
50	12	12	0.1	0.05	32	2580	771	19.90	15.09	21	11

Table 3.4: Results for different asymmetric two dimensional potentials systems simulated without a thermostat. Bias shape parameters (h_H, σ_H) as well as FE barriers ($\Delta G^{\ddagger}_{CC}$, $\Delta G^{\ddagger}_{CH}$) are expressed in kcal/mol, while bias addition frequency δ_H is expressed in ps; masses are expressed in Dalton. Temperature average and standard deviation ($\langle T \rangle$, σ_T) are expressed in K; time average and standard deviation ($\langle t \rangle$, σ_t) are in ps. The "Neg." and "Pos." columns refer to the number of observed simulation ending in the negative ($x < -1$) or the positive ($x > 1$) basins, respectively.

BF	h_H	σ_H	δ_H	a_x	a_y	d_x	Runs	$\langle T \rangle$	σ_T	$\langle t \rangle$	σ_t	Neg.	Pos.
3	0.01	0.1	0.05	2	2	1	60	508	193	5	5	59	1
3	0.05	0.05	1	4	4	1	60	1327	392	2	2	56	4
3	0.05	0.05	1	3	3	1.4	60	940	361	2	2	56	4
∞	0.05	0.05	1	3	3	1.4	60	989	495	2	3	57	3
3	0.05	0.05	1	3	3	1.4	60	488	286	5	4	58	2
3	0.01	0.01	0.1	3	3	1.4	60	1164	424	1	2	56	4
3	0.05	0.05	1	3	3	1.4	60	1252	297	1	2	59	1
3	0.01	0.1	0.05	3	3	0.6	60	587	328	6	3	55	5
25	0.01	0.1	0.05	2	2	0.6	97	1227	603	4	8	76	21
3	0.01	0.1	0.05	2	2	0.6	99	747	245	3	3	88	11
3	0.01	0.1	0.05	10	10	0.6	86	1440	975	7	4	85	1

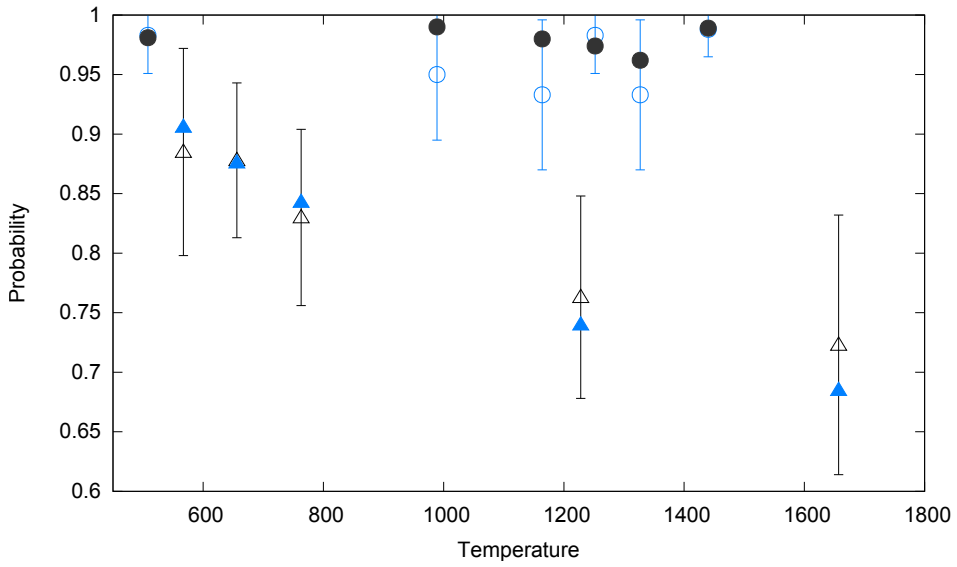


Figure 3.2: Comparison between theoretical (full symbols) and computed (empty symbols) probability for the transition from the central ($-1 < x < 1$) to the negative basin ($x < -1$) for different 2d potentials. Data from simulations performed with (triangles) and without a thermostat (circles); vertical bars show the 95% confidence interval.

3.2 Ethane decomposition

As a second example we considered the first step of unimolecular decomposition of ethane at high temperatures (approximately between 1000 and 3000 K). As with the previous system only two pathways are possible, a C−C or a C−H bond breaking, but this time the reactions have different multiplicities (one and six, respectively) and dissimilar entropic contributions. For all simulations we used adaptive intermolecular reactive bond-ordered potential AIREBO [22]. While classical reactive FF are not necessarily accurate compared to *ab-initio* or density functional theory methods, they provides a consistent and computationally light framework to test our method. Since the method itself is not dependent in any way on the underlying potential, this choice affects the result but not the validity of the tests.

3.2.1 FE calculations

In order to validate the rates computed with FERN we needed an estimate of the rate based on the FE surface at different temperatures for both the possible reactions. The knowledge of the FE difference allows to use Equation 2.3 to compute the theoretical rates under each condition. To obtain the values of $\Delta G_i^{\ddagger}$ at different temperatures, we interpolated the FE computed at 500, 1000, 1500 and 2000 K and assumed the transmission rates equal for all the reactions (see Figures 3.3 and 3.4).

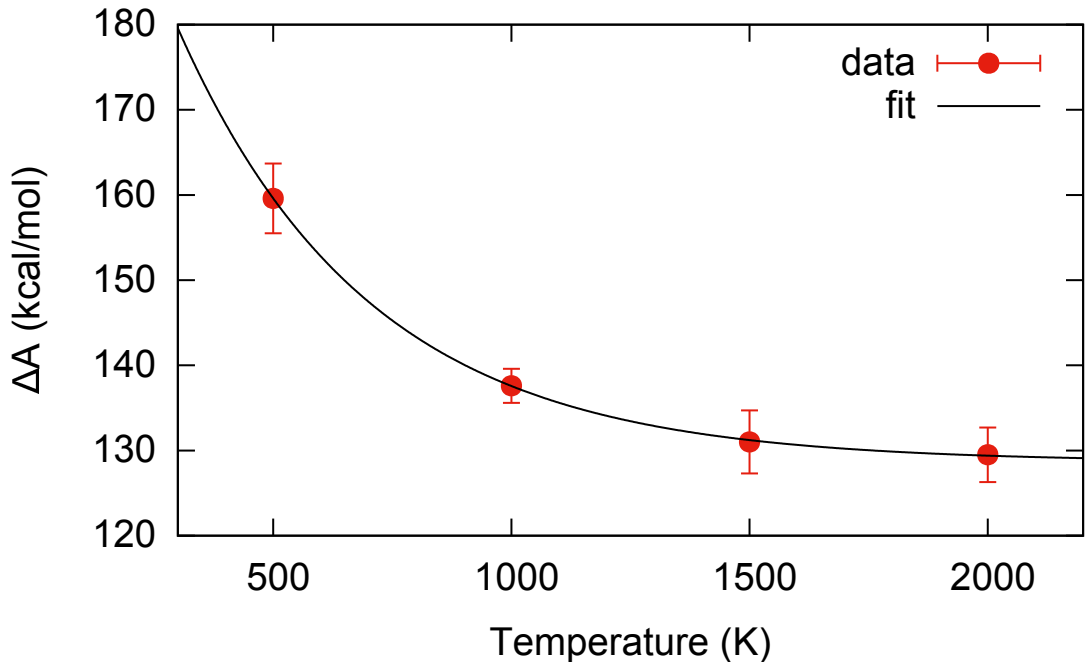


Figure 3.3: Computed FE for the C−C bond breaking in the ethane molecule at different temperatures. Results are shown with the 95% confidence interval. The black line shows the exponential fitting.

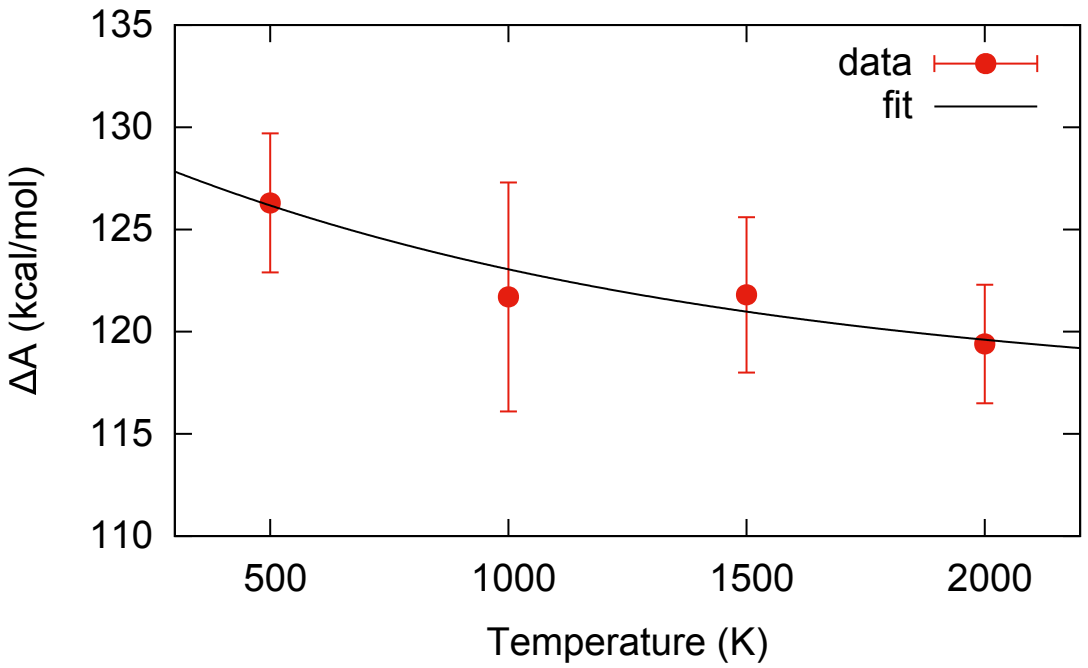


Figure 3.4: Computed FE for the C−H bond breaking in the ethane molecule at different temperatures. Results are shown with the 95% confidence interval. The black line shows the exponential fitting.

The values at the four reference temperature were computed by employing WTM and the relevant bond distance as a CV. For these simulations we employed MD settings similar to the one reported in subsection 2.3.1, with the only exception of the simulation length (150 ns) and the value of the BF which was modified so that $T*$BF was almost constant (about 3500 K).

The data points reported in Figures 3.3 and 3.4 is the FE difference between the bonded and the transition state. The first is defined as the region around the minimum (MIN) that has up to $2k_BT$ more energy than MIN, while the latter is the region around the maximum that has up to $\frac{1}{2}k_BT$ less energy than the barrier.

3.2.2 Results

The FERN method was tested on a variety of different systems, related to both the simulation settings and the physic of the system (isotopic effect) as listed in Table 3.5

We did not apply a thermostat as its efficiency is not constant for all the frequencies and therefore its use can radically influences the results. Instead, the control on the final temperature was obtained by changing the biasing parameters, while still maintaining each added bias relatively small (normally less than 1/10th of k_BT for both Gaussian height and width). Figure 3.5 shows a selection (chose for clarity) of the results, while the complete list is reported in Table 3.6.

As can be seen in Figure 3.5 the predicted probability of a pathway is recovered in a wide range of temperatures and simulations parameters, with a minimal computational cost. Even

Table 3.5: List of the different systems used to test the ethane reactivity. Bias shape parameters (h_H, σ_H) as well as FE barriers ($\Delta G^{\ddagger}_{CC}$, $\Delta G^{\ddagger}_{CH}$) are expressed in kcal/mol, while bias addition frequency δ_H is expressed in ps; masses are in Dalton.

Label	BF	h_H	σ_H	δ_H	m_C	m_H	$\Delta G^{\ddagger}_{CC}$	$\Delta G^{\ddagger}_{CH}$	Runs
EtH1	3	0.01	0.1	0.05	12	1	129	119	100
EtH2	∞	0.01	0.1	0.05	12	1	129	118	200
EtH3	3	0.1	0.25	1	12	1	136	122	39
EtH4	3	0.25	0.25	1	12	1	129	120	40
EtH5	3	0.25	0.25	1	12	1	132	121	40
EtH6	3	1	0.25	1	12	1	129	120	40
EtH7	3	1	0.25	1	12	1	129	120	40
EtH8	3	1	0.25	1	12	1	129	120	38
EtH9	3	1	0.25	1	12	1	129	120	39
EtH10	3	1	0.25	1	12	1	132	121	40
EtH11	3	1	0.25	1	12	1	132	121	34
EtH12	250	1	0.25	1	12	1	129	120	40
EtH13	250	1	0.25	1	12	1	129	120	40
EtH14	250	1	0.25	1	12	1	129	120	40
EtH15	3	4	0.25	1	12	1	133	121	35
EtH16	∞	0.01	0.1	0.05	12	2	129	119	197
EtH17	∞	0.01	0.1	0.05	12	3	129	119	197
EtH18	3	0.01	0.1	0.05	12	4	131	121	70
EtH19	∞	0.01	0.1	0.05	12	4	129	119	194
EtH20	3	1	0.25	1	12	4	129	120	40
EtH21	3	1	0.25	1	12	4	147	124	29
EtH22	∞	0.01	0.1	0.05	12	5	130	120	182
EtH23	∞	0.01	0.1	0.05	12	9	131	121	91
EtH24	3	1	0.25	1	12	9	148	124	11

without any specific optimization, the average simulation time is most of the time 10 ps or less. Considering a few hundred simulations for each system, the total required simulation time is on the order of a few nanoseconds and can be further tuned by modifying the number of simulations performed on each node.

Another interesting result is the conservation of some aspect of the dynamics. Despite not being in any part of the FERN formulation, this feature of can be observed when considering the isotopic effect associated with the change in mass of the H. As $\Delta G^{\ddagger}_{CH}$ is not affected by the change of mass, the number of observed pathway is only minimally affected with respect to the $m_H = 1$ case, but the isotopic effect is recovered in the average reaction time $\langle t \rangle$, which is dominated by the average time for the C-H reaction. Not only the time increases with the mass but also the correct mass dependence ($\propto \sqrt{m_H}$) is recovered as shown in Figure 3.6.

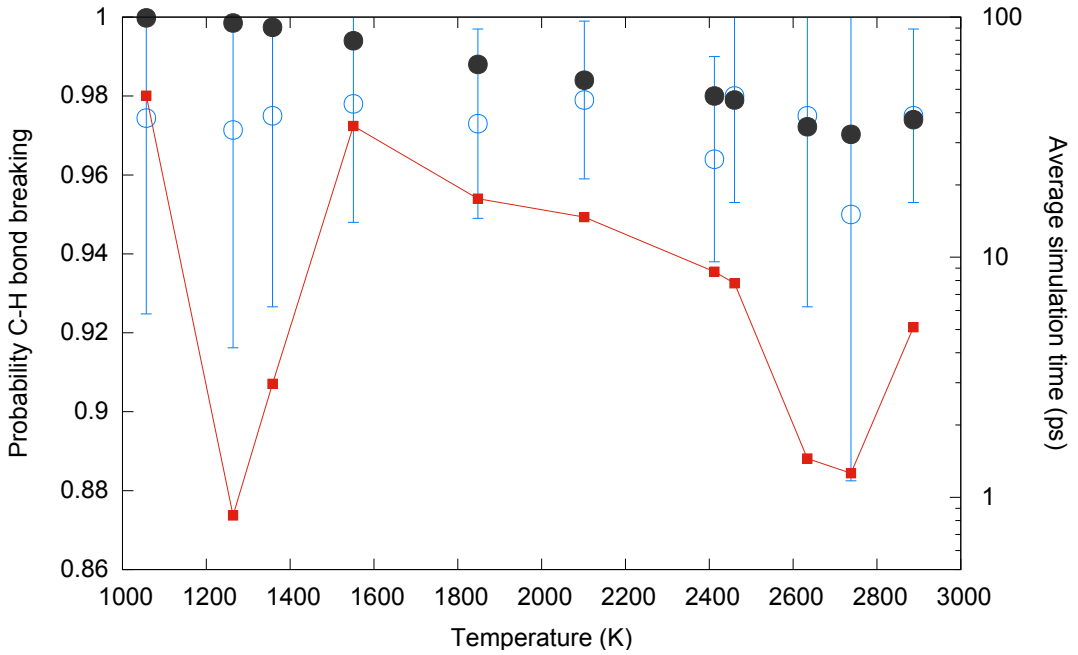


Figure 3.5: (*left y-axis*) Comparison between theoretical (full circles) and computed (empty circles) probability for the C−H bond breaking. Vertical bars show the 95% confidence interval. (*right y-axis*) Squares indicate the average simulation time for each system; each point represents the average of 40 to 200 simulations (see Table 3.5).

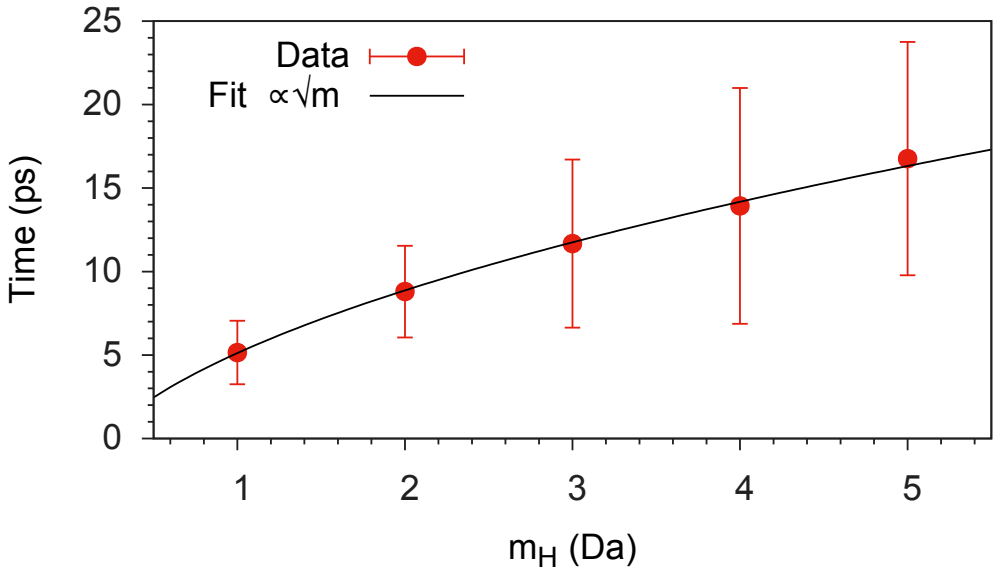


Figure 3.6: Average reaction time for the C−H bond breaking for different H masses. Data with 95 % confidence interval are shown in red; fitting function plotted in black.

Table 3.6: Results for the simulations of the ethane reactivity. Temperature average and standard deviation ($\langle T \rangle$, σ_T) are expressed in K. Time average and standard deviation ($\langle t \rangle$, σ_t) are in ps. The "CC" and "CH" columns refer to the number of observed simulation ending with a C−C or a C−H bond breaking.

Label	$\langle T \rangle$	σ_T	$\langle t \rangle$	σ_t	CC	CH
EtH1	2460	505	7.79	2.14	2	98
EtH2	2887	654	5.11	2.93	5	195
EtH3	1057	374	47.10	31.65	1	38
EtH4	2905	697	3.89	2.16	1	39
EtH5	1417	593	13.05	9.41	0	40
EtH6	4134	970	0.94	0.51	0	40
EtH7	2738	989	1.26	0.92	2	38
EtH8	2632	826	1.25	0.97	0	38
EtH9	2760	1021	1.38	0.94	0	39
EtH10	1358	593	2.97	1.72	1	39
EtH11	1336	720	2.96	1.98	0	34
EtH12	3985	799	0.94	0.49	0	40
EtH13	2634	761	1.45	0.98	1	39
EtH14	2860	994	1.44	1.20	1	39
EtH15	1264	593	0.84	0.78	1	34
EtH16	2412	682	8.70	4.22	7	190
EtH17	2124	663	11.80	7.74	5	192
EtH18	1586	567	23.60	10.09	4	66
EtH19	2102	626	14.70	10.86	4	190
EtH20	2204	646	5.32	2.86	0	40
EtH21	722	176	102.89	68.14	2	27
EtH22	1848	607	17.52	10.75	5	177
EtH23	1551	486	35.28	15.10	2	89
EtH24	701	165	171.73	55.19	0	11

3.3 *t*-decalin + methyl radical

The third application we considered was the reactivity of a *t*-decalin molecule in presence of a methyl radical. The former is used in surrogate fuels and the latter is a common species in combustion environments and plays an important role, among other things, in H abstraction reactions. This system can evolve in eleven different products as shown in Figure 3.7, each one characterized by different multiplicity, barrier height and shape.

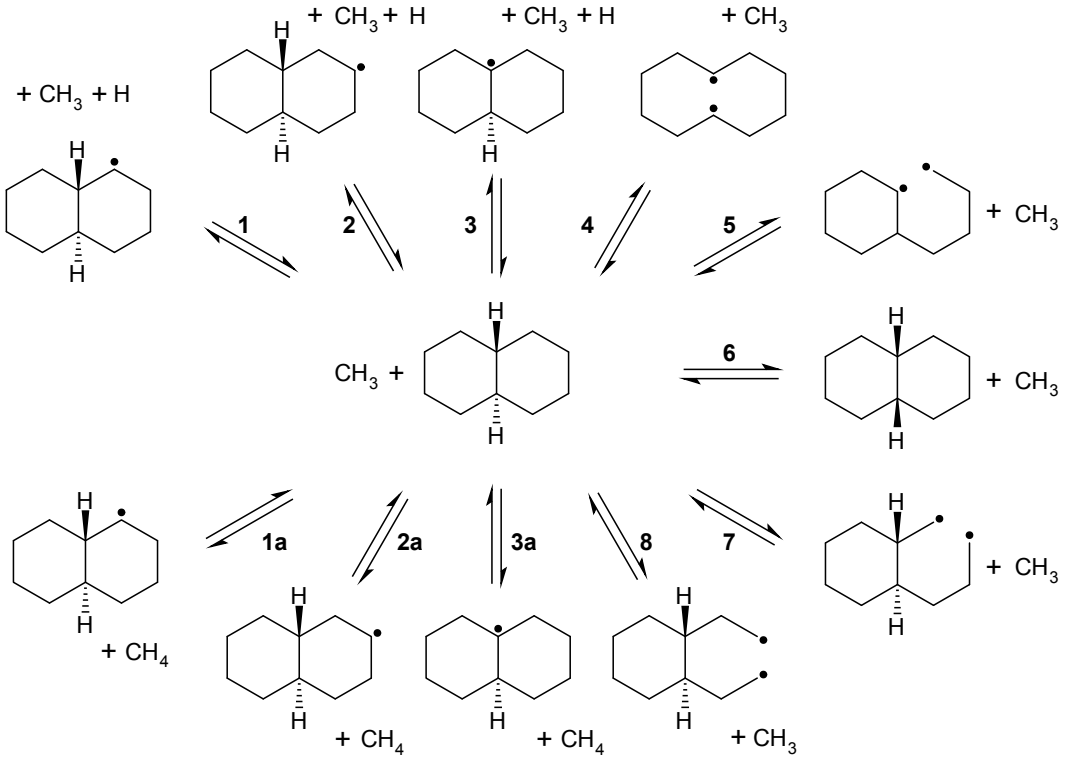


Figure 3.7: Possible reactions of the *t*-decalin + CH_3 system. Six pathways involve the unassisted (**1-3**) and assisted (**1a-3a**) C−H bond breaking, four (**4,5,7,8**) the C−C cleavage, and one (**6**) the isomerization to *c*-decalin.

3.3.1 FE calculations

As before to compare the rate computed with FERN we had to first compute the FE profile associate with each reaction. For this purpose the reactions were separated in 4 groups: the C−H bond breaking (reactions 1, 2, and 3), the C−C bond breaking (reactions 4, 5, 7, and 8), the isomerization (reaction 6) and the hydrogen abstractions (reactions 1a, 2a, and 3a). As before, we computed the FE at five different temperatures (among the set of 500, 700, 750, 800, 1000, 1500 and 2000 K) and then fitted the results with and exponential function. Simulations settings are similar the the one listed for ethane in section 3.2.1. A few examples of the FE profiles for bond breaking reactions are shown in Figures 3.8 and 3.9.

For the first two groups, as shown in Figures 3.10 and 3.11 the differences between each single reactions are in general smaller than the accuracy of the FE calculations. Therefore

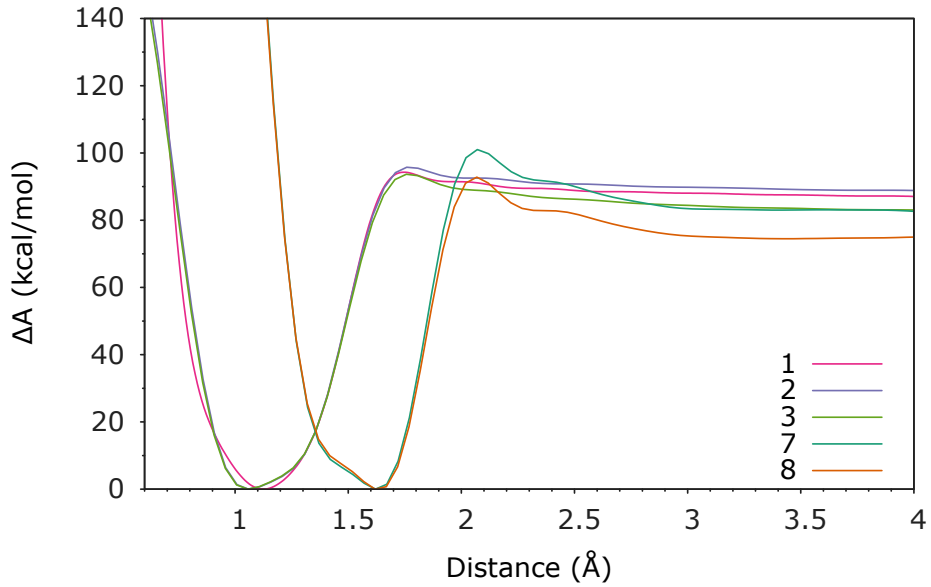


Figure 3.8: Comparison of the FE profiles for selected C−C and C−H bond breaking at 500 K; reactions are labeled according to Figure 3.7.

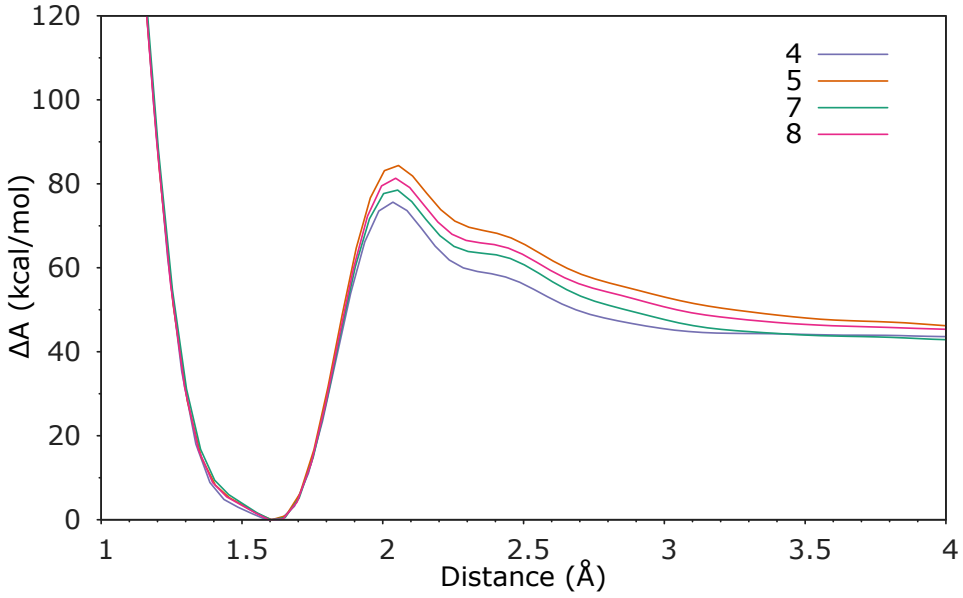


Figure 3.9: Comparison of the FE profiles for C−C bond breaking at 1500 K; reactions are labeled according to Figure 3.7.

we used them together to build a general fitting function by computing for each type of bond breaking.

The hydrogen abstractions were instead studied by simultaneously biasing the specific C−H bond and the distance between the hydrogen and the methyl radical carbon. The results

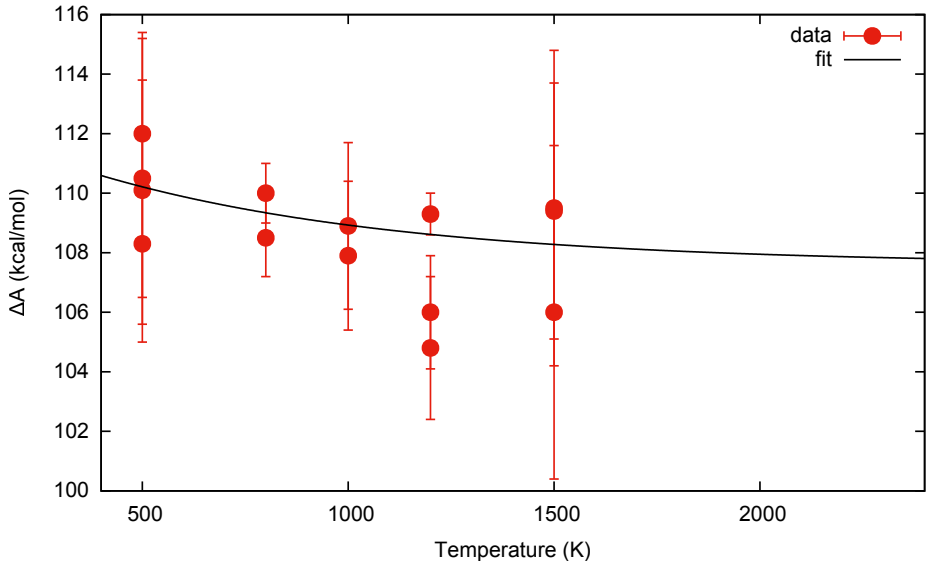


Figure 3.10: Computed FE for the C−C bond breaking for *t*-decalin at different temperatures. Results are shown with the 95% confidence interval. The black line shows the exponential fitting.

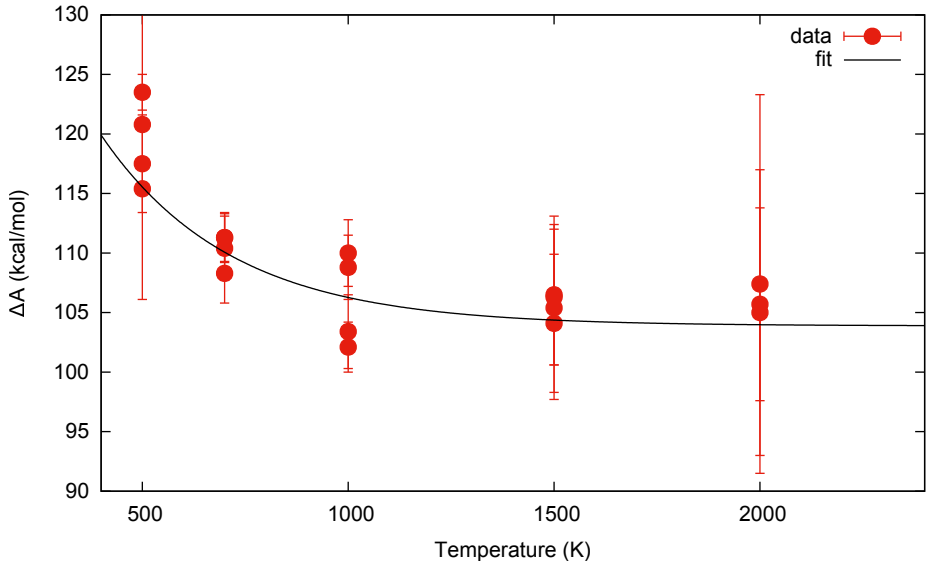


Figure 3.11: Computed FE for the C−H bond breaking for *t*-decalin at different temperatures. Results are shown with the 95% confidence interval. The black line shows the exponential fitting.

of these simulations are reported in Figures 3.12, 3.13, and 3.14.

The surprising result is that there is no communication between the two basins corresponding to the stable states before and after the H abstraction. Therefore, the reactions **1a**, **2a**, and

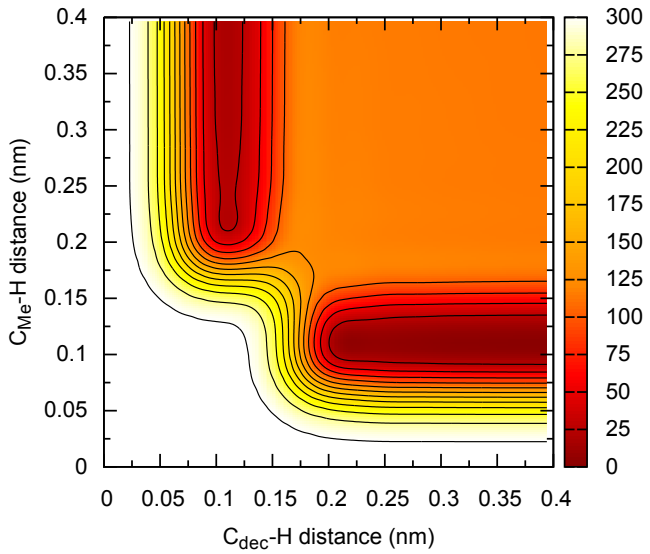


Figure 3.12: FE projections for reaction 1a (Figure 3.7) as a function of the distance of the hydrogen from the C on the decalin and the carbon on the methyl radical. Energy is reported in kcal/mol and isolines are drawn every 25 energy units.

3a are in this FF effectively two step reactions in which the presence of the methyl radical does not facilitate the C−H bond breaking. Alternatively, if we consider them as single step reactions the abstraction must overcome a barrier of at least 50 kcal/mol higher than that needed for the unassisted C−H bond breaking (depending on the assumed reaction path). As a consequence we don't expect to observe these reaction in a sample of a few hundred FERN simulations.

This result is a consequence of the choice of the specific version of the AIREBO FF, which does not change the charge distribution as a function of the distance of the methyl radical. A more recent version of this FF [24] or different reactive potentials may give a more accurate picture of this system's reactivity. Nonetheless, as explained in the methodology section, the validity of our approach is not affected by the choice of the potential and as long as the FERN produces results consistent with the underlying FE, we can consider the technique successful.

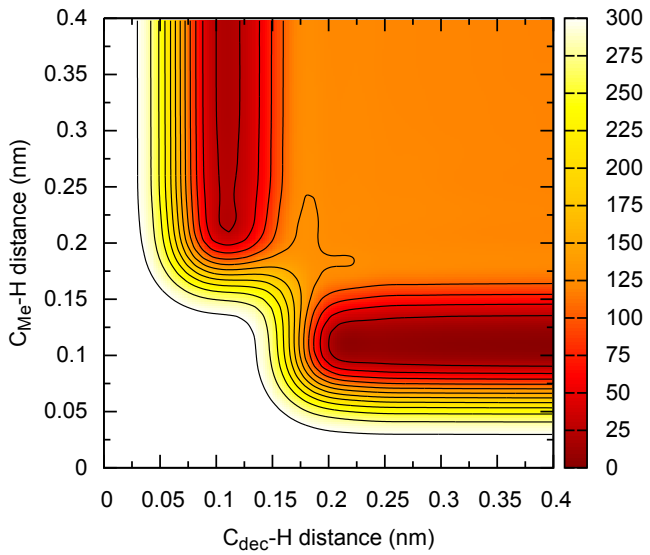


Figure 3.13: FE projections for reaction 2a (Figure 3.7) as a function of the distance of the hydrogen from the C on the decalin and the carbon on the methyl radical. Energy is reported in kcal/mol and isolines are drawn every 25 energy units.

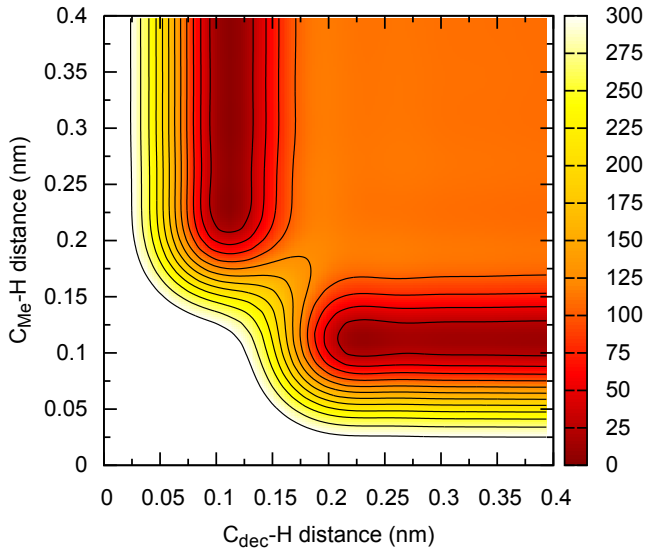


Figure 3.14: FE projections for reaction 3a (Figure 3.7) as a function of the distance of the hydrogen from the C on the decalin and the carbon on the methyl radical. Energy is reported in kcal/mol and isolines are drawn every 25 energy units.

3.3.2 Results

For the FERN simulations of these systems, we followed the same general protocol used for ethane, with the addition of a soft wall placed on the distance between the center of mass of the methyl and the decalin molecules at 0.8 nm. This constraint was added to increase the number of collisions between the radical and the decalin so that the collision frequency would not be a bottleneck in the assisted hydrogen abstraction rates. We tested different conditions as listed in Table 3.7

Table 3.7: List of the different systems used to test the *t*-decalin reactivity. Bias shape parameters (h_H, σ_H) as well as FE barriers ($\Delta G^{\ddagger}$) are expressed in kcal/mol, while bias addition frequency δ_H is expressed in ps. In all cases META algorithm was used (BF $= \infty$). Reactions are marked according to the labels defined in Figure 3.7.

					$\Delta G^{\ddagger}$							
Label	δ_H	σ_H	h_H	Runs	1	2	3	4	5	6	7	8
Set1	0.05	0.1	0.1	181	104	104	104	108	108	115	108	108
Set2	0.01	0.1	0.1	183	104	104	104	108	108	115	108	108
Set3	0.05	0.2	0.2	187	104	104	104	108	108	115	108	108

The results are reported in Table 3.8 and the relative importance of the different reactions rate is compared to the theoretical value in Figure 3.15.

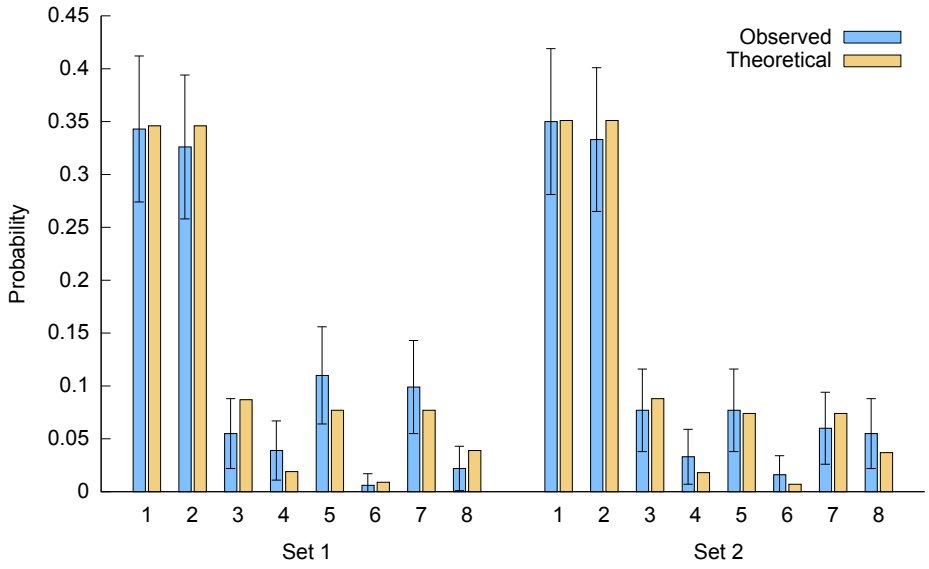
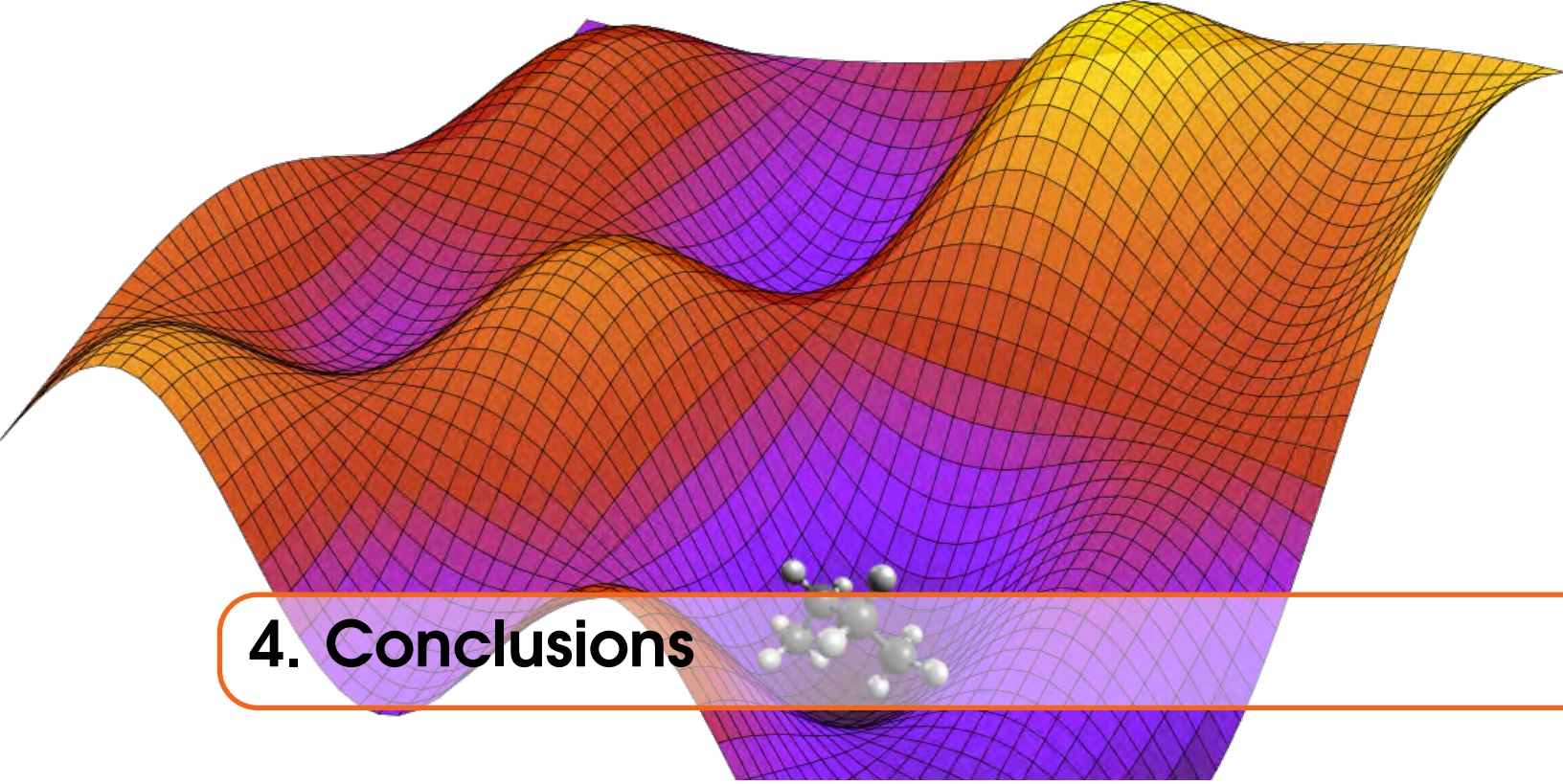


Figure 3.15: Comparison between the observed and theoretical probability of reactions for the *t*-decalin + CH_3 system for two different set of conditions. Reactions are labeled according to Figure 3.7. In both cases the hydrogen abstractions (reactions 1a to 3a) are not reported due their negligible probability. Vertical bars represent the 95% confidence intervals.

Table 3.8: List of the results for the *t*-decalin reactivity tests. Temperature average and standard deviation ($\langle T \rangle$, σ_T) are expressed in K; time average and standard deviation ($\langle t \rangle$, σ_t) are in ps. Reactions are marked according to the labels defined in Figure 3.7.

					Occurrences							
Label	$\langle T \rangle$	σ_T	$\langle t \rangle$	σ_t	1	2	3	4	5	6	7	8
Set1	2424	435	11.52	7.20	62	59	10	7	20	1	18	4
Set2	2246	931	7.40	7.08	64	61	14	6	14	3	11	10
Set3	1928	1418	1.27	1.72	85	65	23	2	6	0	5	1

The results show an excellent agreement for all the pathways and, as before, we observe a tremendous speedup, with average simulation times in a range of 2 to 30 ps depending on the biasing parameters. As expected, in the FERN simulations the hydrogen abstraction reactions are not observed. Since these are indicated in the literature as the dominating pathways in such conditions [25], is clear that the choice of FF plays a crucial role in the determinations of the pathways. In this regard, the substantial speed up of FERN with respect to other methods allows using more accurate techniques than classic MD even with todays computational resources.



4. Conclusions

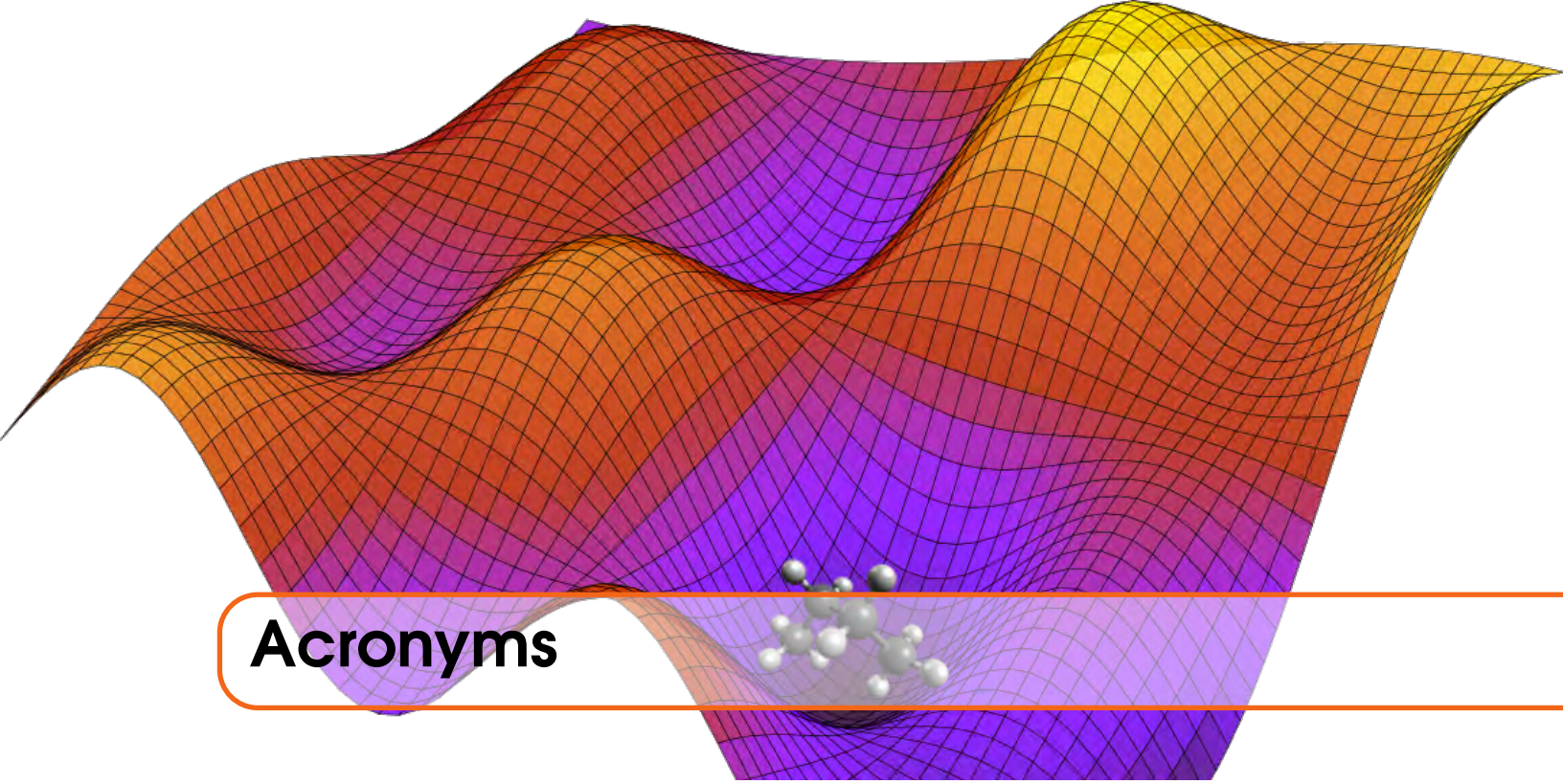
A variety of natural phenomena comprises a huge number of competing reactions and short-lived intermediates. Any study of such processes requires the discovery and accurate modeling of their underlying chemical reaction network. However, this task is challenging due to the complexity in exploring all the possible pathways and the high computational cost in accurately modeling a large number of reactions. Fortunately, very often these processes are dominated by only a limited subset of the network's reaction pathways.

In this work, we propose a novel method with limited computational requirements that is able to identify and select the key pathways of complex reaction networks, so that high-level *ab-initio* calculations can be more efficiently targeted at these critical reactions. The method estimates the relative importance of the reaction pathways for given reactants by analyzing the accelerated evolution of hundreds of replicas of the system and detecting products formation. Within the quite general validity of Equation 1.1 our method identifies the subset of the most likely reaction pathways with a minimal computational effort and without assumptions on the reactions or transition states or the need to define different collective variables for each reaction. Importantly, the method is efficiently iterative, as it can be straightforwardly applied for the most frequently observed products, therefore providing an effective algorithm to identify the key reactions of extended chemical networks. We verified the validity of our approach on three different systems, including the reactivity of *t*-decalin with a methyl radical, and in all cases the expected behavior was recovered within statistical error. These tests, in particular the decalin reactivity, with eleven different pathways that include both unimolecular and bimolecular reactions, show the full potential of our approach: the FERN method is able to tremendously speed up the reactivity of gas phase reactions without requiring any previous knowledge of the system reactivity. The reproducibility of the results independently from

the bias and simulation parameters, as well as the generality of the potential energy/SPRINT combination, makes this method suitable for its effective iterative application to complex reaction networks. Moreover, the options of varying the number of runs for given reactants and blocking specific reactions allow a very efficient exploration even of simple or partially known reaction networks.

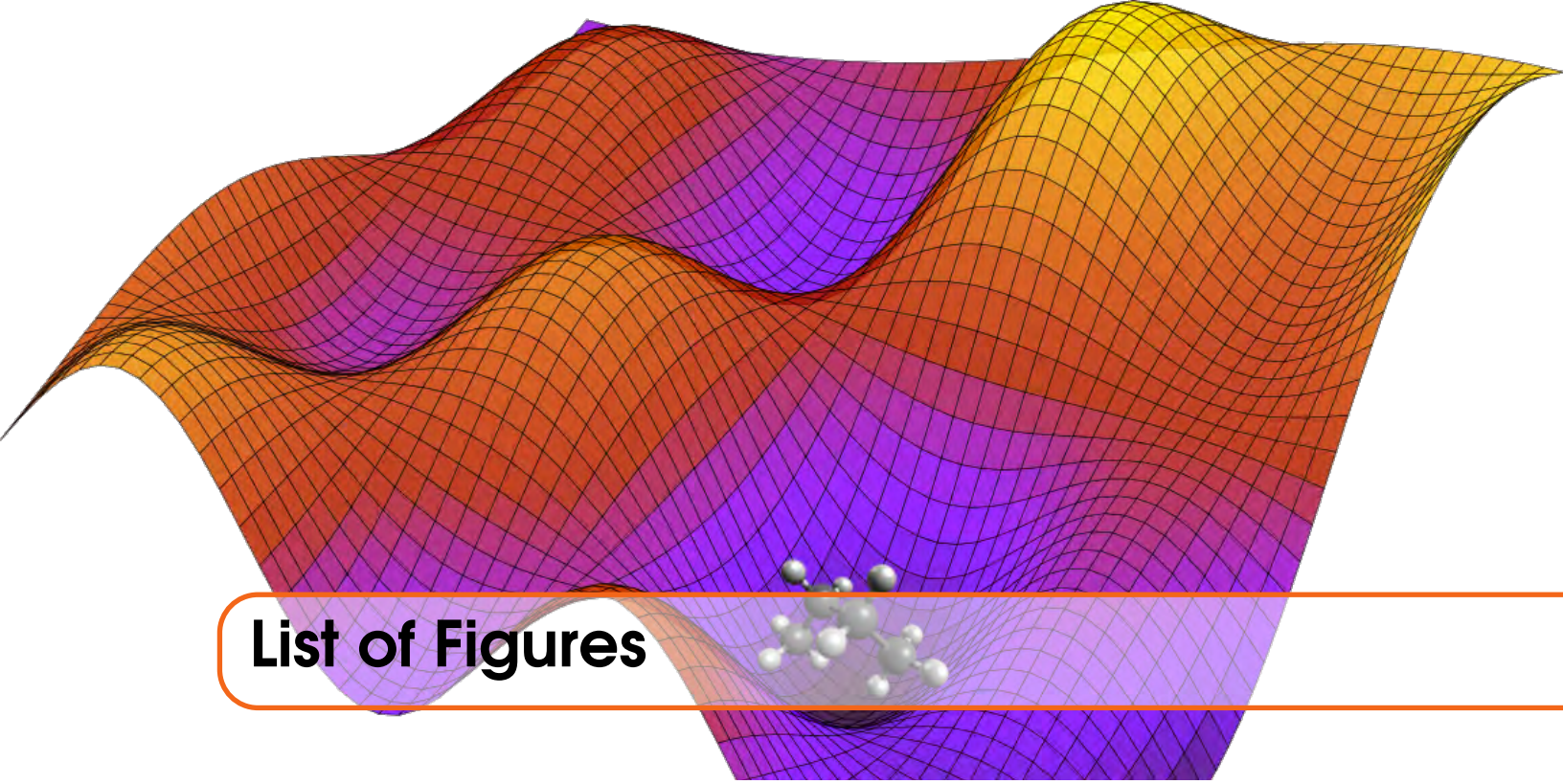
Finally, this acceleration-detection approach can be extended to different classes of systems, with minimal adjustments. For example, while the potential energy and SPRINT combination is a general streamlined choice for systems in the gas phase, for other type of reactive networks, like chemical reactions in solutions, the relevant relaxation times, *e.g.*, water reorientation, should be taken into account so that the behavior of the accelerated system is not biased by a specific initial configuration.

III Additional Material

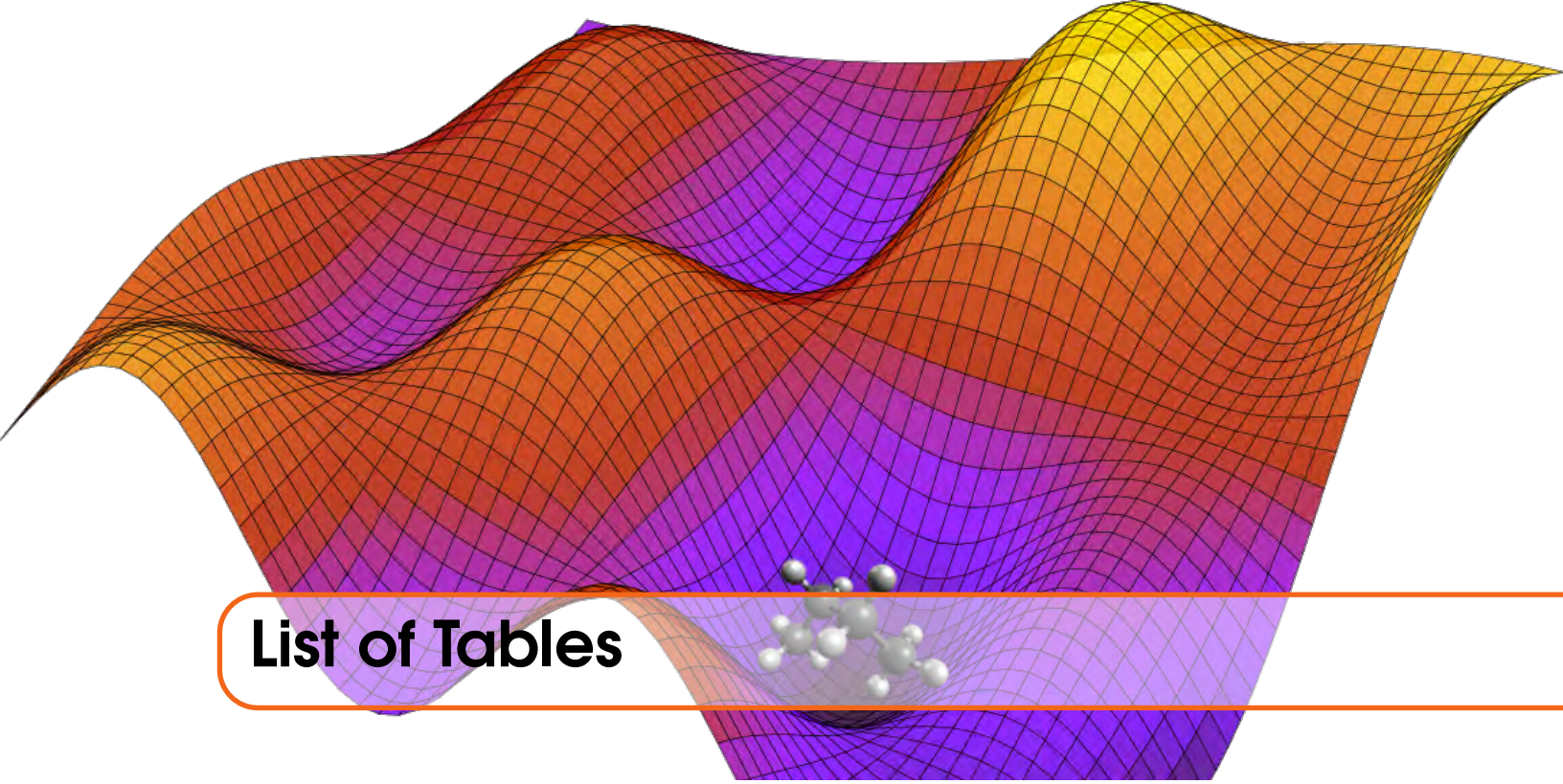


Acronyms

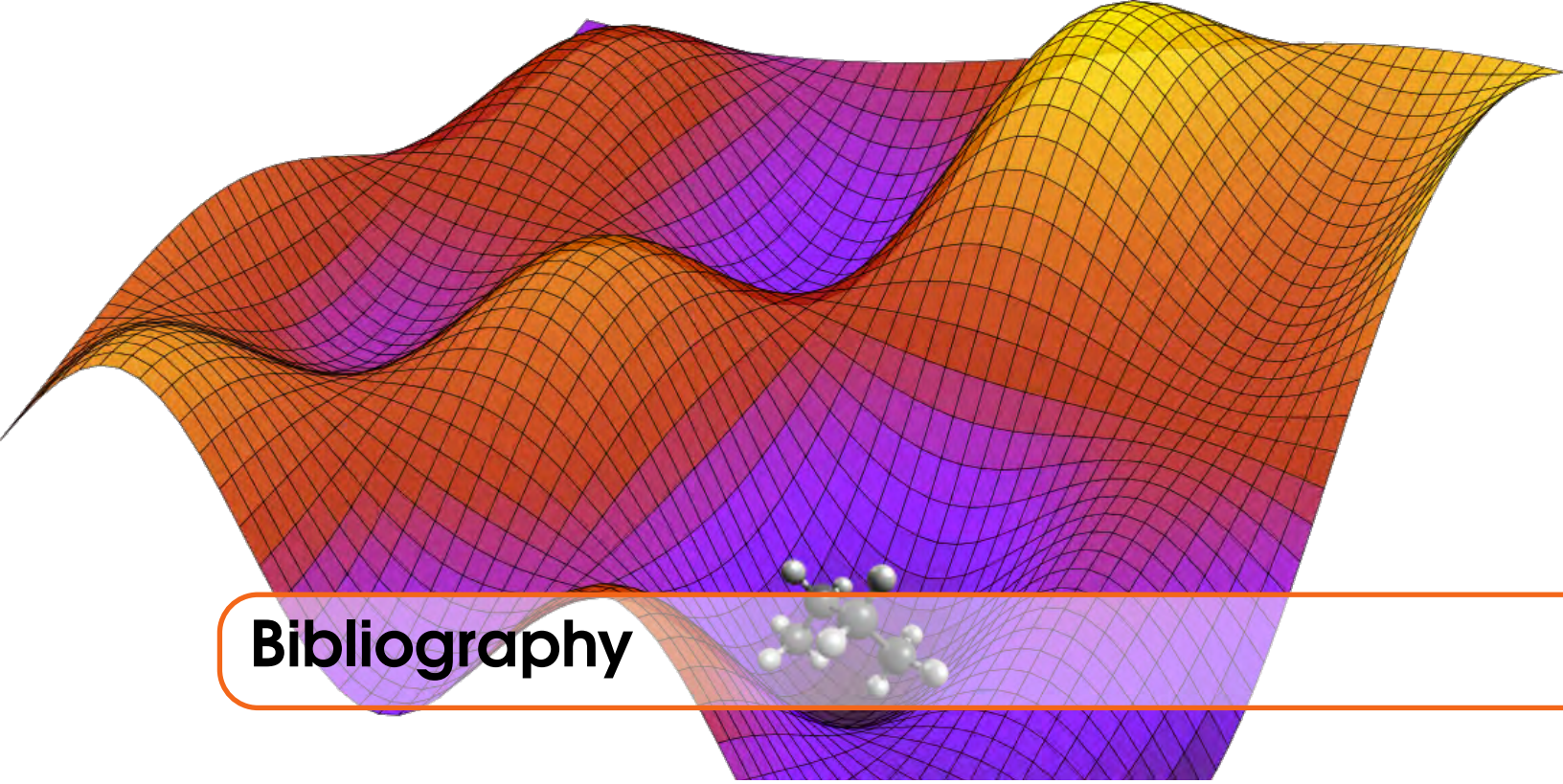
BF	bias factor	6
CN	coordination number	7
CV	collective variable	6
FE	free energy	4
FERN	Fast Exploration of Reaction Network	7
FF	force field	8
MD	molecular dynamics	8
META	Metadynamics	6
RM	reduced mechanism	3
SPRINT	Social PeRmutation INvarianT coordinates	6
WTM	well-tempered Metadynamics	6



List of Figures



List of Tables



Bibliography

[1] David Young. *Computational chemistry: a practical guide for applying techniques to real world problems*. John Wiley & Sons, Apr. 7, 2004. 409 pages. ISBN: 9780471458432 (cited on page 3).

[2] Miles S. Okino and Michael L. Mavrovouniotis. "Simplification of mathematical models of chemical reaction systems". In: *Chemical Reviews* 98.2 (1998), pages 391–408 (cited on page 3).

[3] Alison S. Tomlin and Tamás Turányi. "Mathematical tools for the construction, investigation and reduction of combustion mechanisms". In: *Low-temperature Combustion and Autoignition*. Edited by Michael J. Pilling. Volume 35. Comprehensive Chemical Kinetics. Amsterdam: Elsevier, 1997, pages 293–437. ISBN: 9780444824851 (cited on page 3).

[4] Roberta G. Susnow et al. "Rate-Based Construction of Kinetic Models for Complex Systems". In: *The Journal of Physical Chemistry A* 101.20 (May 1, 1997), pages 3731–3740. DOI: 10.1021/jp9637690 (cited on page 3).

[5] David M. Golden and John R. Barker. "Pressure- and temperature-dependent combustion reactions". In: *Combustion and Flame*. Special Issue on Kinetics 158.4 (Apr. 2011), pages 602–617. DOI: 10.1016/j.combustflame.2010.08.011 (cited on page 3).

[6] Jingzhi Pu, Jiali Gao, and Donald G. Truhlar. "Multidimensional Tunneling, Recrossing, and the Transmission Coefficient for Enzymatic Reactions". In: *Chemical Reviews* 106.8 (Aug. 2006), pages 3140–3169. DOI: 10.1021/cr050308e (cited on page 3).

[7] Donald G. Truhlar, Bruce C. Garrett, and Stephen J. Klippenstein. “Current Status of Transition-State Theory”. In: *The Journal of Physical Chemistry* 100.31 (Jan. 1996), pages 12771–12800. DOI: 10.1021/jp953748q (cited on page 4).

[8] Christoph Dellago et al. “Transition path sampling and the calculation of rate constants”. In: *The Journal of Chemical Physics* 108.5 (Feb. 1, 1998), pages 1964–1977. DOI: 10.1063/1.475562 (cited on pages 4, 5).

[9] Fugao Wang and D. P. Landau. “Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States”. In: *Physical Review Letters* 86.10 (Mar. 5, 2001), pages 2050–2053. DOI: 10.1103/PhysRevLett.86.2050 (cited on pages 4, 5).

[10] Titus S. van Erp, Daniele Moroni, and Peter G. Bolhuis. “A novel path sampling method for the calculation of rate constants”. In: *The Journal of Chemical Physics* 118.17 (May 1, 2003), pages 7762–7774. DOI: 10.1063/1.1562614 (cited on pages 4, 5).

[11] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. “Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method”. In: *Physical Review Letters* 100.2 (Jan. 18, 2008), page 020603. DOI: 10.1103/PhysRevLett.100.020603 (cited on pages 4, 5, 14).

[12] Pratyush Tiwary and Michele Parrinello. “From Metadynamics to Dynamics”. In: *Physical Review Letters* 111.23 (Dec. 3, 2013), page 230602. DOI: 10.1103/PhysRevLett.111.230602 (cited on pages 4, 5).

[13] C. Chipot and A. Pohorille. *Free Energy Calculations*. 1st. Berlin, Heidelberg: Springer, 2007 (cited on page 6).

[14] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (Oct. 1, 2002), pages 12562–12566. DOI: 10.1073/pnas.202427399 (cited on page 6).

[15] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. “Metadynamics”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.5 (Sept. 2011), pages 826–843. DOI: 10.1002/wcms.31 (cited on page 6).

[16] M. Bonomi and M. Parrinello. “Enhanced Sampling in the Well-Tempered Ensemble”. In: *Physical Review Letters* 104.19 (May 10, 2010), page 190601. DOI: 10.1103/PhysRevLett.104.190601 (cited on page 6).

[17] Carine Michel, Alessandro Laio, and Anne Milet. “Tracing the Entropy along a Reactive Pathway: The Energy As a Generalized Reaction Coordinate”. In: *Journal of Chemical Theory and Computation* 5.9 (2009), pages 2193–2196. DOI: 10.1021/ct900177h (cited on page 6).

[18] Fabio Pietrucci and Wanda Andreoni. “Graph Theory Meets Ab Initio Molecular Dynamics: Atomic Structures and Transformations at the Nanoscale”. In: *Physical Review Letters* 107.8 (2011), page 085504. DOI: 10.1103/PhysRevLett.107.085504 (cited on page 6).

[19] Jason Y. W. Lai, Paolo Elvati, and Angela Violi. “Stochastic atomistic simulation of polycyclic aromatic hydrocarbon growth in combustion”. In: *Physical Chemistry Chemical Physics* 16.17 (2014), page 7969. DOI: 10.1039/c4cp00112e (cited on page 7).

[20] Steve Plimpton. “Fast parallel algorithms for short-range molecular dynamics”. In: *Journal of Computational Physics* 117.1 (1995), pages 1–19. DOI: 10.1006/jcph.1995.1039 (cited on page 8).

[21] Massimiliano Bonomi et al. “PLUMED: A portable plugin for free-energy calculations with molecular dynamics”. In: *Computer Physics Communications* 180.10 (Oct. 2009), pages 1961–1972. DOI: 10.1016/j.cpc.2009.05.011 (cited on page 8).

[22] Steven J. Stuart, Alan B. Tutein, and Judith A. Harrison. “A reactive potential for hydrocarbons with intermolecular interactions”. In: *The Journal of Chemical Physics* 112.14 (Apr. 8, 2000), pages 6472–6486. DOI: doi:10.1063/1.481208 (cited on pages 8, 18).

[23] T. Schneider and E. Stoll. “Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions”. In: *Physical Review B* 17.3 (Feb. 1, 1978), pages 1302–1322. DOI: 10.1103/PhysRevB.17.1302 (cited on page 8).

[24] M. Todd Knippenberg et al. “Bond-order potentials with split-charge equilibration: Application to C-, H-, and O-containing systems”. In: *The Journal of Chemical Physics* 136.16 (Apr. 28, 2012), page 164701. DOI: 10.1063/1.4704800 (cited on page 26).

[25] Kyungchan Chae and Angela Violi. “Thermal Decomposition of Decalin: An Ab Initio Study”. In: *The Journal of Organic Chemistry* 72.9 (Apr. 1, 2007), pages 3179–3185. DOI: 10.1021/jo062324x (cited on page 29).

1.

1. Report Type

Final Report

Primary Contact E-mail

Contact email if there is a problem with the report.

avioli@umich.edu

Primary Contact Phone Number

Contact phone number if there is a problem with the report

734-494-0017

Organization / Institution name

University of Michigan

Grant/Contract Title

The full title of the funded effort.

A New Paradigm to Identify Reaction Pathways in Gas-phase

Grant/Contract Number

AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-13-1-0031

Principal Investigator Name

The full name of the principal investigator on the grant or contract.

Angela Violi

Program Manager

The AFOSR Program Manager currently assigned to the award

Chiping Li

Reporting Period Start Date

02/01/2013

Reporting Period End Date

01/31/2015

Abstract

The complexity of the energy landscapes of hydrocarbon molecules during combustion processes, often composed by several hundreds of minima and even more barriers, makes the heuristic search of the most likely reactions an unfeasible task even with today's computer power. As a consequence, the main advancements in the field of combustin chemistry and kinetic mechanisms development are based on "chemical intuition" and trial and error procedures, which are error prone and hard to automate. In this proposal we present a new paradigm to determine reaction pathways for gas-phase species that relies on two major components: molecular dynamics simulations and advanced sampling techniques. Molecular dynamics (MD) in conjunction with advanced sampling techniques, such as Metadynamics, are used to explore the energy landscapes of uni-molecular and bi-molecular reactions in gas phase. Starting from the fuel molecule and running several simulations from the various wells identified with MD, we can recover a network of reactions that includes a controlled number of reactions, allowing the construction of reaction pathways with the desired level of detail.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

SF298_Violi_April2015.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

Violi_Final Report_AFOSR 2015.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

Elvati, P., Violi, A., 2014. “Fast exploration of chemical reaction networks.” arXiv:1405.6610 [cond-mat, physics:physics]

Changes in research objectives (if any):

Not applicable

Change in AFOSR Program Manager, if any:

Not applicable

Extensions granted or milestones slipped, if any:

Not applicable

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user